

# SEARCH REQUEST FORM

4/26/82  
120

Requestor's  
Name

Rita Mitra

Serial

Number

09-556246

Date

4-25-82

Phone

608-1211

Art Unit

1653

Rm 9803

MB 9801

## Search Topic

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search

AA1

Seg 1, aa 1-24 plus aa 200-1404 (delete aa 25-199)

AA2

Seg 1, aa 1-25 plus 67-1404 (delete aa 26-66)

AA3

Seg 1, aa 1-25 plus aa 67-106 plus aa 200-1404  
(delete aa 26-66 and 107-199)

aa 200-1140 of Seg 1

Seg 1 aa 25-end (delete aa 1-24)

AA4

Seg 1 aa 1-66 plus 105-end (delete aa 67-104)

AA5

Seg 1, aa 1-156 plus aa 200-1404 (delete aa 157-199)

AA6

Seg 1, aa 1-106 plus aa 200-1404 (delete aa 107-199)

AA7

Seg 1, aa 200-1167

AA8

Seg 1, aa 200-1212

AA9

Seg 1, aa 200-1263

Point of Contact

Toby Port

Technical Info Specialist

CM16A04

703-308-3634

## STAFF USE ONLY

Date completed

4/29

Searcher

T. Port + B. Brynner

Terminal time

120

Elapsed time

20

CPU time

Total time

Number of Searches

Number of Databases

Search Site

SMC

CM-1

Pre-S

Type of Search

N.A. Sequence

A.A. Sequence

Structure

Bibliographic

Vendors

IG Suite

STN

Dialog

APS

Geninfo

SDC

DARC/Questel

Other

T6C+26

BEST AVAILABLE COPY













"Preoperative diagnosis of thyroid papillary carcinoma by reverse transcriptase polymerase chain reaction of the MUC1 gene.";

[11]  
SEQUENCE OF 1-89 FROM N.A.  
TISSUE=Lung;  
RX MEDLINE=96181716; PubMed=8604237;  
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,  
Lee L.N., Luh K.T., Wu C.W.;  
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and  
tissues.";  
RL Oncology 53:118-126(1996).

[12]  
SEQUENCE OF 1-46 FROM N.A.  
RC TISSUE=Breast carcinoma;  
RA Buluwela L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;  
RL Submitted (Oct-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CYTOSKELETON.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM  
IS ALSO PRODUCED.  
CC -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE  
SPLICING.  
CC -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL  
TUMORS, SUCH AS BREAST CANCER.  
CC -!- PTM: HIGHLY GLYCOSYLATED (N- AND O-LINKED CARBOHYDRATES AND SIALIC  
ACID).  
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT  
VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE  
MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.  
CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.

-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; J05582; AAA60019.1; -  
EMBL; M32738; AAA35804.1; -  
EMBL; M32739; AAA35806.1; -  
EMBL; J05581; AAA59876.1; -  
EMBL; M61170; AAB53150.1; -  
EMBL; X52229; CAA36478.1; ALT\_SEQ.  
EMBL; X52228; CAA36477.1; ALT\_SEQ.  
EMBL; M35093; AAB59612.1; ALT\_SEQ.  
EMBL; Z17324; CAA78972.1; -  
EMBL; Z17325; CAA78973.1; -  
EMBL; M31823; AAA35757.1; -  
EMBL; S81781; AAD14376.1; ALT\_INIT.  
EMBL; S81736; AAD14369.1; ALT\_INIT.  
EMBL; M21868; AAA59874.1; ALT\_SEQ.  
PIR; B35175; B35175.  
PIR; B35175; B35175.  
PIR; S10218; S10218.  
GlycoSuiteDB; P15941; -  
MIM; 158340; -  
MIM; 113720; -  
InterPro; IPR000082; SEA.  
Pfam; PF01390; SEA; 1.  
SMART; SM00200; SEA; 1.  
PROSITE; PS50024; SEA; 1.  
Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
Repeat; Alternative splicing.  
SIGNAL 1 23 POTENTIAL.  
CHAIN 24 1255 MUCIN 1.  
FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1163 1186 POTENTIAL.  
FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 181 960 44 X 20 AA TANDEM REPEATS.  
FT DOMAIN 1034 1151 SEA.

FT	CARBOHYD	957	957	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	975	975	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1029	1029	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1055	1055	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1133	1133	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARSPPLIC	19	19	T -> TATTAPKAT (IN ISOFORM B).	
FT	VARSPPLIC	20	22	MISSING (IN ISOFORM C).	
FT	VARSPPLIC	20	31	MISSING (IN ISOFORM D).	
FT	VARSPPLIC	126	905	MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).	
FT	VARSPPLIC	1077	1087	FLQYKGGFL -> VSIGLSFFMLP (IN SECRETED ISOFORM).	
FT	VARSPPLIC	1088	1255	MISSING (IN SECRETED ISOFORM).	
FT	CONFLICT	2	2	T -> A (IN REF. 11).	
FT	CONFLICT	134	134	P -> Q (IN REF. 9).	
FT	CONFLICT	154	154	P -> Q (IN REF. 9).	
FT	CONFLICT	1021	1021	S -> T (IN REF. 3).	
FT	CONFLICT	1251	1251	A -> T (IN REF. 3).	
SQ	SEQUENCE	1255 AA;	122072 MW;	5E28DFC4DE7D9A82 CRC64;	

Query Match 8.9%; Score 651; DB 1; Length 1255;  
Best Local Similarity 27.8%; Pred. No. 9.3e-19;  
Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;

QY	236	TSLVNKETTVETKETTNNKQSTDKKETS	SAKETQSIKTS	SAKD-LAPTSKVLAKPT	294		
Db	16	TVLTV-----VTGSGHASSTPGGKET	SATORSVSPS	TEKNAVMTSSVLSHS	65		
QY	295	P-KAETTKGP--ALITPKP-----	-----P	TTPKPEASTTPKEPTTTI---K	334		
Db	66	PGSGSSTQGDVTLAPATEPAGS	AAATWQDVTS	VEVTRPALGSTTPPAHDVTS	125		
QY	335	SAPTTKPEAPTTKSAPTTKPEAPTTK	PE-----	APTTPKEAPTTKPEAPTTTK	388		
Db	126	PAPGSTAPPAGVT-SAPDT--RPAG	STAPPAGVTS	SAPDT--RPAGSTAPPAGVT--	179		
QY	389	SAPTTKPEAPTT-----	-----	KKPAITPKP-----APTTPKEPTTTP	426		
Db	180	SAPDTRPAGSTAPPAGVTSAPDTR	PAAGSTAPP	AGVTSAPDTRPAGSTAPPAGVT	239		
QY	427	-----KEPAPTTKPE-----	APTTPKEAP	TPAK-----KPAPTTKPEAPTTK	467		
Db	240	SAPDTRPAGSTAPPAGVTSAPDTR	PAAGSTAPP	AGVTSAPDTRPAGSTAPPAGVT	299		
QY	468	PAPTTTKESPTTPKEPAPTT-----	TKSAPTTK	PEAPTTTKSAPTTTKESPTTPKEP	522		
Db	300	SAPDTRPAGSTAPPAGVTSAPDTR	PAAGSTAPP	AGVTSAPDTRPAGSTAPPAGVT	356		
QY	523	-----APTTPKEAPTTK-----	-----	KPAPTTKPEAPTTK-----EPAPTTTKKAP	570		
Db	357	GVTSAPDTRPAGSTAPPAGVTSAPDTR	PAAGSTAPP	AGVTSAPDTRPAGSTAPPAGVT	416		
QY	571	APTTPKETAPTTKLTPTTPEKLAPTT	PEAPTTPE	ELAPTTPEAPTTPEAPTTPEAPTT	630		
Db	417	GVTSAPDTRPAGSTAPPAGVTSAPDTR	PAAGSTAPP	AGVTSAPDTRPAGSTAPPAGVT	472		
QY	631	PKA-----AAPNTPKPEAPTTK-----	EPAPTTK	PEAPTTKPEAPTTKPEAPTTKPEAPTT	682		
Db	473	PPAHGVTSAPDTRPAGSTAPPAGVTS	SAPDTRRPA	GGSTAP-----PAHGVTSAPDTRRAP	528		
QY	683	APTTP-----	-----	KKPAKELAPTTTKETSTTS	SDKPAP--TTPKGTAPTTTKEPAPT	730	
Db	529	GSTAPPAGHVTSAPDTRPAGSTAP	PAHGVTS	SAPDTRPAGSTAPPAGHVTS-----	APD	583	
QY	731	TPKEAPTTPKG-----	TAPTTLKE	PAPTTP-----	-----	KKPAKELAPTTTKGTSTT	776
Db	584	TRPAGSTAPPAGHVTSAPDTRPAG	STAPPAG	HVTSAPDTRPAGSTAPPAGSTAP	PAHGVTSAP	642	
QY	777	SDKPAP--TTPK-----	ETAPTTK	PEAPTT-----	-----	KKPAITTPETTPPTTSVS	820
Db	643	DTRPAGSTAPPAGHVTSAPDTRPAG	STAPPAG	HVTSAPDTRPAGSTAPPAGHVTS	-----	TAPPAGHVTS	700

Db 588 TTSSSAPVPT----PSSSTSSSAPAPTPSSSTSSSAPVPT----SSSTSSSAP-V 638  
QY 599 PEKPAPTTPEELAPT-TPEEPTTTPPEPAP-----TTPKAAAPNTPEKPAPTTPEKAP 652  
Db 639 PTPSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVPT----SSSTSSSAP 694  
QY 653 TTPKPAPTTTPKETAPT-TPKGTAPTTLKEPAPTTPKPAKELAPTTPKPTTSTTSKDP 711  
Db 695 VT-----SSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVPTPS-SSSTSS 749  
QY 712 APTTPKGTAPTTPKEPAP---TTPKEPAPT-TPKGTAPTTLKEPAPTTPKPAKELAPT 767  
Db 750 APVTSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPV 809  
QY 768 TTKGPTSTSDKPAPTTPKETAPTTPKEPAP-----TTPKPAKELAPTTPKPTTSEVSTP 822  
Db 810 PTPSSSNITSSAPSTPSSSTSSSAPVPTPSSSTSSSAPVSSSTSSSAPVPTP 869  
QY 823 T-----TTKEPTTIHKSPDESPELSAETPKALENSPKPEGVPTTKTPAAATKPEMTT 877  
Db 870 SSSSNITSSAPSI---PFSSTTESFSTGT-TVTPSSSKYPGSGTETSVSSTETIVPT 925  
QY 878 KDKTTERDLRTTPEETTAAPKMKETATTEKTESKITAT--TQVISTTTTQDTPPEK 935  
Db 926 KTTSTVPTSTTTTITTVCTGNSAGETSGCSPKTVTTVPTTTTTSVTTSSITTT 985  
QY 936 TLLKTTTLAPKVTT---TKKTIITT-EIMNKPEETAAPKDRATNSKATPKPKQK----- 986  
Db 986 TVCSTGNSAGETSGCSPKTIITTVPCSTSPSETA-----SESTTSTPTPTTVTV 1037  
QY 987 -----TKAPKKPTSTKKPTMDRVRKPTTPPKMTSTMPLENPSTRIAEAM 1034  
Db 1038 STTVVTTEYSTKPGGEITTFVKNIPITTYLTITAPD--SVTVTNETPTT-ITTV 1094  
QY 1035 LQTTTRPNQPNKLVNPKSDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQ 1094  
Db 1095 CSYGT-----NSAGETSGCSPKTVTTVPCSTGTGEVTEATTLVTRAVTT 1141  
QY 1095 IINPMISDETNCNGPVDGLTLNGLTVLAFRGHYFWMLSFPSPSPARRITEVWGIP 1154  
Db 1142 TVVTTESSTGTNSA-GKTTTGYTTKSVPTT-----YVTLAPSAVPTPATN-----AVP 1189  
QY 1155 SPIDTVETRC-----NCEGKT 1170  
Db 1190 TTITT--TECSAATNAAGET 1207  
  
RESULT 4  
MUC1\_HUMAN  
ID MUC1\_HUMAN STANDARD; PRT: 1255 AA.  
AC P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEMT)  
DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)  
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-  
DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN  
DE DF3).  
GN MUC1.  
DN MUC1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=90368716; PubMed=2394722;  
RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;  
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";  
RL J. Biol. Chem. 265:15294-15299(1990).  
[2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=90202794; PubMed=23118825;  
RA Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;  
RT "Epistatin, a carcinoma-associated mucin, is generated by a  
RT polymorphic gene encoding splice variants with alternative amino  
RL termini.";  
RL J. Biol. Chem. 265:5573-5578(1990).  
[3]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Breast carcinoma;  
RC MEDLINE=90368713; PubMed=1697589;  
RX Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,  
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;  
RT "Molecular cloning and expression of human tumor-associated  
RT polymorphic epithelial mucin.";  
RL J. Biol. Chem. 265:15286-15293(1990).  
[4]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91097524; PubMed=2268309;  
RX Lancaster C.A., Peat N., Duhig T., Wilson D.,  
RA Taylor-Papadimitriou J., Gendler S.J.;  
RT "Structure and expression of the human polymorphic epithelial mucin  
RT gene: an expressed VNTR unit.";  
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).  
[5]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Breast carcinoma;  
RC MEDLINE=90276413; PubMed=2351132;  
RX Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,  
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,  
RA Keydar I.;  
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing  
RT may generate multiple protein forms.";  
RL Eur. J. Biochem. 189:463-473(1990).  
[6]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Breast carcinoma;  
RC MEDLINE=90276414; PubMed=2112460;  
RX Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,  
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;  
RT "A transcribed gene, containing a variable number of tandem repeats,  
RT codes for a human epithelial tumor antigen. cDNA cloning, expression  
RT of the transfected gene and over-expression in breast cancer  
RT tissue.";  
RL Eur. J. Biochem. 189:475-486(1990).  
[7]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91033045; PubMed=1688329;  
RX Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,  
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;  
RT "Isolation and characterization of an expressed hypervariable gene  
RT coding for a breast-cancer-associated antigen.";  
RL Gene 93:313-318(1990).  
[8]  
RN PARTIAL SEQUENCE FROM N.A.  
RP MEDLINE=88330762; PubMed=3417635;  
RX Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,  
RA Burchell J.;  
RT "A highly immunogenic region of a human polymorphic epithelial mucin  
RT expressed by carcinomas is made up of tandem repeats.";  
RL J. Biol. Chem. 263:12820-12823(1988).  
[9]  
RN SEQUENCE OF 1-169 FROM N.A.  
RP MEDLINE=90088473; PubMed=2597151;  
RX Abe M., Siddiqui J., Kufe D.;  
RT "Sequence analysis of the 5' region of the human DF3 breast  
RT carcinoma-associated antigen gene.";  
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).  
[10]  
RN SEQUENCE OF 1-109 FROM N.A.  
RP TISSUE=thyroid;  
RX MEDLINE=96183746; PubMed=8608966;  
RX Weiss M., Baruch A., Keydar I., Wreschner D.H.;





This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

```

CC      EMBL; L21998; AA95295.1; -
CC      EMBL; M74027; AA59875.1; -
CC      EMBL; M94131; AA59163.1; -
CC      EMBL; M94132; AA59164.1; -
CC      MIM; 158370; -
CC      InterPro; IPR000359; Cys_knot.
CC      InterPro; IPR000561; EGF-like.
CC      InterPro; IPR002400; GF_cys_knot.
CC      InterPro; IPR001007; VWFC.
CC      InterPro; IPR001846; Wd.
CC      Pfam; PF00007; Cys_knot; 1.
CC      Pfam; PF00094; vwd; 4.
CC      PRINTS; PR00438; GFCYSKNOT.
CC      SMART; SM00214; VWC; 2.
CC      SMART; SM00011; VWC_def; 2.
CC      SMART; SM00216; VWD; 4.
CC      PROSITE; PS00022; EGF_1; UNKNOWN_1.
CC      PROSITE; PS01185; CTCK_1; 1.
CC      PROSITE; PS01225; CTCK_2; 1.
CC      PROSITE; PS01208; VWFC; 2.
CC      Glycoprotein; Repeat; Signal.
FT      SIGNAL      1      20      POTENTIAL.
FT      CHAIN       21      5179
FT      DOMAIN     1401      1747      MUCIN 2.
FT      REPEAT     1401      1416      APPROXIMATE REPEATS.
FT      REPEAT     1417      1432      1.
FT      REPEAT     1433      1448      2.
FT      REPEAT     1449      1464      3.
FT      REPEAT     1465      1471      4.
FT      REPEAT     1472      1478      5.
FT      REPEAT     1479      1494      6.
FT      REPEAT     1495      1517      7A.
FT      REPEAT     1518      1533      7B.
FT      REPEAT     1534      1556      8A.
FT      REPEAT     1557      1572      8B.
FT      REPEAT     1573      1596      9A.
FT      REPEAT     1597      1612      9B.
FT      REPEAT     1613      1635      10A.
FT      REPEAT     1636      1651      10B.
FT      REPEAT     1652      1675      11A.
FT      REPEAT     1676      1683      11B.
FT      REPEAT     1684      1699      12.
FT      REPEAT     1700      1715      13.
FT      REPEAT     1716      1731      14.
FT      REPEAT     1732      1747      15.
FT      REPEAT     1748      1763      16.
FT      DOMAIN     4815      4886      VWFC 1.
FT      DOMAIN     4924      4991      VWFC 2.
FT      DOMAIN     5075      5160      CTCK.
FT      DISULFID   5075      5122      BY SIMILARITY.
FT      DISULFID   5089      5136      BY SIMILARITY.
FT      DISULFID   5098      5152      BY SIMILARITY.
FT      DISULFID   5102      5154      BY SIMILARITY.
FT      DISULFID   ?        5159      BY SIMILARITY.
FT      CARBOHYD   163      163      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   423      423      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   670      670      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   770      770      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   894      894      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1139      1139      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1154      1154      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1215      1215      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1230      1230      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1246      1246      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1787      1787      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   1820      1820      N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT      CARBOHYD   4339      4339      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4351      4351      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4362      4362      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4373      4373      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4422      4422      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4438      4438      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4502      4502      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4616      4616      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4627      4627      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4752      4752      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4787      4787      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4881      4881      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4888      4888      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4955      4955      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   4970      4970      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   5019      5019      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   5038      5038      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD   5069      5069      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CONFLICT   1351      1351      H -> L (IN REF. 3).
FT      CONFLICT   1412      1412      T -> S (IN REF. 3).
FT      CONFLICT   1449      1449      L -> P (IN REF. 3).
FT      CONFLICT   1504      1504      M -> T (IN REF. 3).
FT      CONFLICT   4192      4192      G -> S (IN REF. 2).
SQ      SEQUENCE   5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;

Query Match      16.1%; Score 1171; DB 1; Length 5179;
Best Local Similarity 29.2%; Pred. No. 3.1e-38;
Matches 399; Conservative 79; Mismatches 524; Indels 364; Gaps 52;

QY      29      SCAGRCGEYSRDATCNCYNCOHYMEC-----CPDKRVCCTAAEVHNPTS 74
DB      1079      SCS--CDTG--GDCEFCSAVASAQECTKEGACVFWRTDLCPIF-----CDYNNPH 1128

QY      75      PPSKKAPPPGASQTIKSTT-----KRSPK--PPNKKTKKVISEE--- 115
DB      1129      ECEWHYPCGNRSFETCRITNGIHSNISVSYLGGCYPCPKDRPIYEEDLKKCVTADKCG 1188

QY      116      --ITEEH---SVSENQSSSSSSSSSSSTI-----WKIKSSKNS 150
DB      1189      CYVEDTHYPGASVPTTEETKSCVCTNSQVVCBEGKILNQTOGAPCYWEI-----C 1243

QY      151      AANRELQK-----KLKVDKNKNETKKKPKPPV----- 181
DB      1244      GPNGTVEKHNICSTIRPSTLTITFTTLPTPTSTTTTPTSTTSTLSTLTKLCC 1303

QY      182      -----DEAGSLDNGDFK-----VTPDTSTTQH-NKVSTSPK--- 213
DB      1304      LMSDWINEDHPSSGSDGSDREPFDGVCAPEDIECRSVKDPHLSLEHQKQVCDVSYGF 1363

QY      214      -----ITTAKPINRPRSLPPNSDTSKETSILVN 241
DB      1364      ICKNEQDFGNGPFGLCYDYKIRVNCWPMDCIITTPSPPTTTTPSPPTTTTLPTTTPS 1423

QY      242      KETTVEKTTTTNKQSTSDGKEKTTSAKETQSIKTSKADLAPTSKVLAETPKAETTT 301
DB      1424      PPTT-----TTTTPTTPTTTPSPPTTTTP-----LPTT-----TPSPITSTT 1462

QY      302      KGPALTTPKEPT-----PTTPKEPASTTPKEPTPTTIKSAK-TPPKPE-----APTPT 348
DB      1463      PPTPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPPTTTPPTT 1522

QY      349      KSAP--TPPKEPATTTKEPAPTTTPKEP-----APTTPKEPAPTTTKSAP--TTPKEPA 398
DB      1523      PSPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1582

QY      399      PTTKKPAPTTTPKEPAPTTTPKEPTTTPKEPAPTTTPKEPAPTTTPKEP-----APTAPKK 451
DB      1583      TTTSPPTTITTTTPTTPTTTPSPPTTTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1642

QY      452      PAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 508
DB      1643      PPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 1702

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:37 ; Search time 62.75 Seconds  
(without alignments)  
798.154 Million cell updates/sec

Title: AA6  
Perfect score: 7294  
Sequence: 1 MAWKTLPIYLLLSVEFIQ.....ARAITRSQTLKSVVYNCP 1366

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues 100059  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1171	16.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.0	1664	1 SLPI_CLOTH	Q06852 clostridium
3	810.5	11.1	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	8.9	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	559	7.7	875	1 FPL_MYTCD	Q25460 mytilus edu
6	555.5	7.6	2700	1 ZAN_MOUSE	Q9Y493 homo sapien
7	551	7.6	620	1 EXTN_TORAC	P13983 nicotiana t
8	551	7.6	1087	1 NFH_MOUSE	P19246 mus musculus
9	530.5	7.3	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.3	865	1 CPN_DROME	Q02910 drosophila
11	518.5	7.1	872	1 FPL_MYTGO	Q25434 mytilus cor
12	517.5	7.1	831	1 NFH_RAT	P16884 rattus norv
13	509.5	7.0	662	1 MUC1_XENLA	P13983 nicotiana t
14	503.5	6.9	1020	1 NFH_HUMAN	P19246 mus musculus
15	502	6.8	1970	1 RPB1_MOUSE	P23253 trypanosoma
16	498.5	6.8	1970	1 RPB1_MOUSE	Q02910 drosophila
17	493.5	6.8	467	1 BAT2_HUMAN	Q25434 mytilus cor
18	490	6.7	2142	1 SSP2_PLAYO	P16884 rattus norv
19	488.5	6.7	826	1 EXTN_MAZE	Q05049 xenopus lae
20	475.5	6.5	267	1 ZAN_MOUSE	P12036 homo sapien
21	471.5	6.5	5376	1 ZAN_MOUSE	P08775 mus musculus
22	468.5	6.4	634	1 HWP1_CANAL	P11414 cricetus
23	467	6.4	817	1 VRP1_YEAST	Q01443 plasmodium
24	454	6.2	797	1 VGLX_HSVB	P46593 candida alb
25	454	6.2	1161	1 YJ9P_YEAST	P37370 saccharomyc
26	453.5	6.2	1229	1 N121_HUMAN	P28968 equine herp
27	448.5	6.1	670	1 V650_HSV11	P47179 saccharomyc
28	444.5	6.1	2774	1 MAPA_RAT	Q092n3 homo sapien
29	443.5	6.1	1794	1 YAVI_SCHPO	Q00130 ictaluria h
30	442.5	6.1	751	1 FPL_MYTGA	P34926 rattus norv
31	439.5	6.0	1083	1 T2D3_HUMAN	Q10172 schizosach
32	436.5	6.0	1411	1 TCOF_HUMAN	Q27409 mytilus gal
33	435.5	6.0	3256	1 K167_HUMAN	Q00268 homo sapien
					Q13428 homo sapien
					P46013 homo sapien

34	432.5	5.9	439	1 XP2_XENLA	P17437 xenopus lae
35	432	5.9	3164	1 TEGU_HSV11	P10220 herpes simp
36	428	5.9	2517	1 NCR2_HUMAN	O9Y618 h nuclear r
37	427.5	5.9	1251	1 YQUS3_CAEEL	Q09550 caenorhabdi
38	426	5.8	2715	1 TRX2_HUMAN	Q9um66 homo sapien
39	425.5	5.8	1185	1 DRPL_HUMAN	P54259 homo sapien
40	424.5	5.8	1125	1 MAP4_MOUSE	P27546 mus musculu
41	424	5.8	307	1 SGS3_DROME	P02840 drosophila
42	423.5	5.8	3421	1 TEGU_HSVB	P28955 equine herp
43	422.5	5.8	1183	1 DRPL_RAT	P24258 rattus norv
44	419.5	5.8	2476	1 ZAN_FIG	O28983 sus scrofa
45	411.5	5.6	907	1 VGP3_EBV	P03200 epstein-bar

ALIGNMENTS

RESULT	ID	MUC2_HUMAN	STANDARD:	PRT:	5179 AA.
AC	Q02817	Q14878			
DT	01-JUN-1994	(Rel. 29, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMUC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
[1]					
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Intestine;				
RC	MEDLINE=94132002; PubMed=8300571;				
RX	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RA	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	"Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor.";				
RL	J. Biol. Chem. 269:2440-2446(1994).				
[2]					
RN	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RP	TISSUE=Colon;				
RC	MEDLINE=93016075; PubMed=1400449;				
RX	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,				
RA	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region.";				
RL	J. Biol. Chem. 267:21375-21383(1992).				
[3]					
RN	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RP	MEDLINE=91358717; PubMed=1885763;				
RX	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RA	Petersen G.M., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RT	and polymorphism.";				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	- SUBUNIT: MULTIMERIC.				
CC	- SUBCELLULAR LOCATION: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	VARIES AMONG DIFFERENT ALLELES.				
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKWORM HEMOCYTIN.				
CC	- SIMILARITY: CONTAINS 2 VWFC DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).				





Search completed: April 26, 2002, 16:32:20  
Job time: 735 sec

Db 2417 SKOTISSTPTDITTAASEEPTKSTMSDLSSTSNVLSSTSTPSESS-----KSPVSSST 2472  
QY 647 -----PKEPAPT-----TKEPAP-----TTKETAPTTPKGTAPTTLKE 681  
Db 2473 EGISVSTSTEFKVPSESTISSVLEEDLTKTPSPILLETITASETSEPLTSLTVSVRI 2532  
QY 682 PATTTPKKPAPKELAPTTTKEPTSTSDKPA-----PT-----T 715  
Db 2533 HELTTSSENVKSESTSTSSSESKSPQBPAGILLTSTVVVPTSSVSLITASEIAITSNT 2592  
QY 716 P--KGAP--TTK-----EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKEL 764  
Db 2593 PFKQGRPTTSPKSLVKSTSESTVTSSEPSSESTKRITVSTVSTTTTEETTTSESLLI 2652  
QY 765 APTTTKGT--STTSRDKPAPTTPKETAPTTPKEPAPTTPKKPAPTTPETPPPTTSEVSPT 823  
Db 2653 LTAAPSKPTSTESSEAPTTPAKTSETKPSNVSTSRKSTENVTSTSQSGLSESSYMS 2712  
QY 824 TTKEPTTIHKSDESTP--ELSAEPTPKALE--NSGKEPGVPTTKTTPAATK-----PEMT 874  
Db 2713 STS-----SEPTNAPAVTVSSEASTTLEENSTSP-----TSSEASVLSLFLPESI 2762  
QY 875 TTAOKDITRDLRTPTTTPAAPTTPKKTATTTTETKSTKITATT----- 920  
Db 2763 TSEAVTVSSR-----APAETMSSESHRELSTVSSEPSPEIPLSTVSPNVVTASSIPSE 2818  
QY 921 -QVSTTTTODTTP-FKITP-----LKTTLAPKVTTTKTITTTIMNKPETAKPKDRA 973  
Db 2819 EPILSVSTSSPTVRLLITGTDDLLIVTVVPSHGNRQNTASSV---PNSNSTPIILP 2875  
QY 974 TNSKATTPKQKPTKAPKPTSTKKPKTWPVRKPK-----TTPTPKMTSTPELN----- 1025  
Db 2876 SES-UTTPQPPPTTTTAKPATTSGKRGPPSIQPPAEMFTTAP-----PPPSNGGYGE 2928  
QY 1026 PFSIAEAMLOTTT 1039  
Db 2929 ETNOEEQVSTTT 2942

RESULT 14  
151618  
nucleolar phosphoprotein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999  
C:Accession: I51618; S57757  
R:Calrns, C.; McStay, B.  
J:Cell Sci. 108, 3339-3347, 1995  
A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180  
A:Reference number: I51618; MUID:96019267  
A:Accession: I51618  
A:Status: preliminary;  
A:Molecule type: mRNA  
A:Residues: 1-990 <CAI>  
A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CAA61368.1; PID:g895921  
C:Genetics:  
A:Gene: xNopp180  
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein  
C:Keywords: phosphoprotein

Query Match 8.8%; Score 643.5; DB 2; Length 990;  
Best Local Similarity 24.2%; Pred. No. 2e-20;  
Matches 246; Conservative 148; Mismatches 424; Indels 199; Gaps 40;

QY 74 SPSSSKAPPSPG-----ASOTIKSTTKRSPKPNKKKT-----KKVISEEITBE 119  
Db 57 SPDAKKRPANGLPPLPKKSKESSESSSEDEEPPAKKRAQAPAGGKPPVKAQPKKA 116  
QY 120 HVSSENOESSSSSSSSSTIWK-ISKSSKNSAANRELOKLLKLVKDNKNRT-----KK 172  
Db 117 KSSSESDSDSEETTKPKPAKRAQTPKVAVKTPQKKAKSSSESSSSSEDEASKKK 176  
QY 173 KPTPKPP---VDEAGSGLDNGDFKVTPTDPTSTTHKNVSTSPKITTAKPINRPSLPN 229

Db 177 QPVIKVPKQAVVAGLASNNG-----KTADSSSSSDSPPAKKTAAATK-----TTP- 224  
QY 230 SDTSKETSLSLVNKETTVETKTTTNTKOTSDGKEKTTSAKETOSIEKTSAKDLAPTS-- 287  
Db 225 --TKPATAAKPOAKTAGKSKSSREDDSSDEEQKTAISKPKPDVYSAPV---PPTSVS 279  
QY 288 --KVLAKTPPAETTTKGPALTTTPKEPTTPKE--PASTTPKEPTTPPTTIKASPTTKEP 343  
Db 280 KKKTLSPQGTAKPESDSSDSSDEEQPAKKAKIVPAKAAASAPKPLAKAETSTDSSES 339  
QY 344 APTT-----TKSAPTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 391  
Db 340 DSSSEDEKSSVKLVGAAP--KKAPA-----APOAKSTPVAAKKSAKPAKASSSD 390  
QY 392 TTPKEPAPTTPKPKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 450  
Db 391 SDSSSNEETTKPAKATTPAKSA-ATPTSKTPTNGK-ATPTSKTTPAKPGTPTSTAKKDS 448  
QY 451 KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 510  
Db 449 SSSDSSDSSDEETTKPAKATTPAKSAATPTSKTPTNSKATPTSKTTPAKPGTPTSA 508  
QY 511 PKESPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 569  
Db 509 KKDSSSDSSDSSDEKKTFA---KRAAKTTPAKPAKTTAKPA-----AKTTPAK 557  
QY 570 PAPTTPKETAPTTPKLTPTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 629  
Db 558 PA-----AKSTPKQVPTPKE-----SSSDSSDSSSEDEKKS 590  
QY 630 TPKAAANTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 689  
Db 591 SAKPAVKTT---PKATSK---PVVASKVPV-AKKAASSSDSSDSEETTKTKPLTKLS 643  
QY 690 PAPKELAPTTTKEPTSTSD-----KPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPK 742  
Db 644 PAVKTLPPKKAESSDSSDSEKKTTPAKPAKATPVNTKAPAKONKASKASDSDS 703  
QY 743 TAPTTLKEPAPTTPKPKAPKELAPTTTGTSTSDKAPTTPKAPTTPKAPTTPKAPTTPK 802  
Db 704 SSEEKSKQP-TGKSPAATAKAP-PKKNPVAVNKDKPSSSSSSDSSGDEKQ---KPK 757  
QY 803 KPAPTTPETPPPTTSEVSPTTTPKEPTTHKSPDESTPELSAEPPTKALENSPKEPG 862  
Db 758 QAA-----AAKDVKQGAKAAPKTPKKAASSSE---DS 787  
QY 863 TKTPAAKPEMTTAKDKTTERDLRTTPTTTPETTTAAAPKMTKETATTTETKTESKITAT 922  
Db 788 SDEEDVSKAKTNTAVSKSPV---TTPKAVPAKKESSSESSDSEDEKQGGKNTSTTKI 843  
QY 923 TSTTTQDTPPKITTLTKTTLAPKVTTTK--KTITTEIMNKPETA-KPKDRATNS--- 976  
Db 844 ANST-----PKAAAECSSESSSESSDEGKANGTSGKRRKRESTGNAEC 885  
QY 977 KATTPKQKPTKAPKKTPTTKPKTTPVRKPKTTP-TPKMTSTPELNPTSRITAE 1032  
Db 886 EAVTPE-----NKKLKAQSPNTFPKVNKELANTFPFRRVVEEDIEINP---RMAD 932

RESULT 15  
T11622

extensin class 1 precursor - cowpea  
C:Species: Vigna unguiculata (cowpea)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 17-Nov-2000  
C:Accession: T11622; S54155  
R:Arsejanjevic-Maksimovic, I.; Broughton, W.J.; Krause, A.  
Mol. Plant Microbe Interact. 10, 95-101, 1997  
A:Title: Rhizobia modulate root-hair-specific expression of extensin genes.  
A:Reference number: 217301; MUID:97155574  
A:Accession: T11622  
A:Status: preliminary;  
A:Molecule type: DNA

```
QY 478 -----PTTKEPAPTTTKSAPTTT--KEPAPTTT 504
Db 2286 GLVCRNEQVGKCMFNEIRVFCNCYCHCPTSTATSPSTPGTTWLLTLTTTA 2345
QY 505 KSAPTTPKPSPTTKEPAPTTTPKPAPTTPKEPAPTTTPKPAPTTPKKA- 563
Db 2346 TTTTSTGATSPSTGCPAGTTPH--VSTTATPTVTSSK---ATPSSPGTATAPAL 2399
QY 564 -----PTAPKEPA-----PTTPKETAPTTPKKLTP----- 588
Db 2400 RSTATTPTATSFATPSSSLGTTWRLSQTTPMATMSTATPSSPETVHTSTVLTTTAT 2459
QY 589 -----TTPEKLAPTTPEKAPTT-----PELAPTTPEEP-----TPTTPEEPAP 628
Db 2460 TTGATGSVATPSPSTPGTAHTTKVPTTTTGTFTVPSSSPGTATPPVWISITTTTPTSGS 2519
QY 629 TTPKAAAPNTPKPAP-TTPKEPAPT--TPKEPAPT--TPKEPAPT--TAPTTPKGTAPTTL 679
Db 2520 TVTPSSIPGTTHTTVLTTTQTVATGSMATPSSSQTSTGTPPSLITATTITATGTT- 2578
QY 680 KEPAPTPPKKPAKELAPT-TTKPEPTS-----TTSDKPAPTTPKGT-----APT 723
Db 2579 -NPSSTPGTPIPELTTTATTAATSSVTVPSSALGTTHTPPVPNTTATTHGRSLSPSS 2637
QY 724 P-----KEPAPTPKEPAPTPKGTAPT-TLKEPAP-----TTPKKP 759
Db 2638 PHTVRTAWTSATGTLGTHTEPSTGTSHTPAATGTTTTPALSSHPSSRTTESPP 2697
QY 760 AKELAPTTTKGPTSTSDKPAPTTPKETAPT--PKP----- 796
Db 2698 SP-----GTTTPGHTATSRITATATPSTKTRTSLLPSTQPTAPITTVTTGCEPQCAWSE 2753
QY 797 ----- 796
Db 2754 WLDYSYMPGPGSGGDFDYSNIRAAAGAVCEQPLGLECRATAOPGVPLGELQGVVECSLD 2813
QY 797 -----APTTPKPAPTTPPETPTT-----SEVSTP 822
Db 2814 FGLVCRNEQVGKCMFNEIRVFCNCYCHCPTSTATSPSTPGTTWLLTLTTT 2873
QY 823 TTTKEPTTIHKSPPDETPELSAEPTPKALENSKPEGVPTK-----TP 866
Db 2874 ATTTASTGSTATP-SSTP-GTAPPPKVLTSPTATPTATSSKATSSSPRTATTLPLVLS 2930
QY 867 AATK-----PEMTTAKDKTTERDLRTTPETT-TAAPKMTKETAT 905
Db 2931 TATKSTATSVTPSSSLGTTGLTPEQTTTPVATMSIHPSSTPETHTSTVLTKATT 2990
QY 906 TTEKT-----TESKITATTQVTS-TTQDTP-----FKITTLKTTTLAP 945
Db 2991 ATSSSTPSPSTPGTTWLLTLTTTAAATTTAGTGATPSTPGTTWLLTLTTTAST 3050
QY 946 KVTTP-KKITITTEIMNKEE---TAKPKDRAT--NSKATTPKQ-----KPKAPKK 992
Db 3051 GSTATLSSTPGTTWLLTEPSTAVTAPPGSTATASQTQATAGTTPHVSSTATTITVTSSK 3110
QY 993 PSTKKPKT---MPRVKPKTTPPKMT-----ST 1020
Db 3111 ATPSSPGTATAPALRSTATTPTATSFATPSSSLGTTWRLSQTTPPTATMTSTATPSS 3170
QY 1021 MPELNPTSRIBAMLOTT-----TRPNQTPNSKLIVENPKSEDAG--GAECETPHMLLR 1072
Db 3171 TPEVHTSTVLTTTATTGATGSVATPSPSTPGTAHTTKVPTTTTGTATPSSSGTALT 3230
QY 1073 PHVEMPEVTPDMDYLPRVFNQIIINPMLSDETNICNGKVPDGLTTLRNGTLVAFRGHYF 1132
Db 3231 PPWILSTTTPTTTP--TTSGSTVTPSSIPTGTTHTARVLTTTITTTTATGSM----- 3280
QY 1133 WMLSPFSPSPARRIEVWGIFSPIDTVETRCNCEGT 1170
Db 3281 -----ATPSSS---TQTSGTPPSLTTTATTITATGST 3309
```

RESULT 13  
T34513  
hypothetical protein ZK783.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34513  
R:Favella, A.; Vaudin, M.  
submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid ZK783.  
A:Reference number: 221536  
A:Accession: T34513  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3507 <FAV>  
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
A:Experimental source: strain Bristol N2; clone ZK783  
C:Genetics:  
A:Gene: CESP:ZK783.1  
A:Map position: 3  
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 8.9%; Score 647.5; DB 2; Length 3507;  
Best Local Similarity 23.3%; Pred. No. 4.9e-20;  
Matches 283; Conservative 182; Mismatches 496; Indels 253; Gaps 52;

```
QY 27 LSSCAGRCGEGSRD-ATC-----NCDYNCOHYMECCPDFKRVCTAAAEVHNP 73
Db 1781 MGSCGCKMAGYTGDCATCIKEEPPKDKTAC--TDEWSRLCELEKKQCTVDEEVPQ 1837
QY 74 SPSSKAPPPGASQTIKSTTKRSPKPPNKKTKKVIIESEI-TEEH----- 120
Db 1838 CGACLPHPHPINGCQSLOI-----SGLCAQKNDCKNKAECIDIHPSHFCPCPDGFIG 1893
QY 121 SVSENESSSSSSSSSSSTIWIKSKNSAANRELQKLVKDKNKNKTKKTPKPPV 180
Db 1894 MICDDVDECNAGCDDENT--KCENTIGFNCVCLGKFKVDEKCVVDEKKQPNREKIE 1951
QY 181 VDEAGSLDN-GDFKVTT---PDTSTQHNKVSTSPKLTTA-----KPINRPSLPP 228
Db 1952 IDEENSSSSNSGQEKPTTKGIVSSTSATSESTTAEAPHVTTTSSSTTKDMTSSKS-PE 2010
QY 229 N---SDTSKETSITVKNKETT-ETKETTTNKOTSTDGKEKTTSAKETOSIEKTSARDLA 284
Db 2011 NVTMSSEPEVSSSSSSSTTASETTIVSSPSSSSSEAPLTSSPATTEVITESSVKSTT 2070
QY 285 P---TSKVLAKTPPKAETTTKGPALTTPKEPTTPKPEASTTPKEPTTTIKS-APT 339
Db 2071 PKESSSEITVKLSSKSPVETSSVKSSPSTPS-TTTSQSVTSTVPETSKSVLSSSEAPVT 2129
QY 340 PKEPAPTTTTSAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 399
Db 2130 STSPTEVHT--SSETKPSLSASSTTGDNSTTPTSSSLASVKTSAPGTSASVAPVKLSS 2188
QY 400 TTPKKPAPTTPK-EPAPTTTPKEPTTPTPKPEAPTTKEP-APTTPKPEAPTPAPKKPAPTTP 457
Db 2189 LSPDVQSPTKPTPDATSESSVQASETSSTGTSVKSTSEPESHVTKLSITSSNPSSSVPT 2248
QY 458 KEPAPTTKPEAPTTTKEPPTTP-KEPAPTTTTSAPTTTKEPAPTTTTSAP-----T 509
Db 2249 PKSTPTVPE-----STEQPTSTTPSGSLTPMNSNSEVLTTPSEPHVLUSSSLDVSOSST 2303
QY 510 TPKEPSTTTTKEPAPTTTPKKKAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTT----- 558
Db 2304 TPNNLSESSIVE---TPKTSSEVSLNSEESTTEAPTLSPDILSTTTNNLSQSSTVST 2359
QY 559 -----TKKPAPTAPK-----EPAPTTKETAPTTPKLTPPTPEKLAPTT 598
Db 2360 EDRSEISSENSEKPT-SAPELVTSVTHVASSSPDVPTES--SEPDDLTSSTENIPEAS 2416
QY 599 PEKAPATTPEELAPTTPEEPTTPPEEP-----APTTPKAAPNTPKPEAPTT- 646
```



Biochem. Biophys. Res. Commun. 155, 644-649, 1989

A>Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated

A:Reference number: A36735; MUID:90088473

A:Accession: A36735

A:Molecule type: mRNA

A:Residues: 1142,'O',144-162,'O',164-168 <ABE>

A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543

R:Masuzawa, Y.; Miyachi, T.; Hamaoune, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.

J. Biochem. 112, 609-615, 1992

A>Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu

A:Reference number: JX0235; MUID:93123189

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>

A:Experimental source: gastric carcinoma cell

R:Zrihan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.

FEBS Lett. 356, 130-136, 1994

A>Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine

A:Reference number: S51026; MUID:95080414

A:Contents: annotation

A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region an

C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c

C:Comment: partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly

C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48162

C:Genetics:

A:Gene: GDB:MUC1; PUM

A:Cross-references: GDB:120705; OMIM:158340

A:Map position: lq21-lq23

A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3

C:Superfamily: polymorphic epithelial mucin

C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis

F:1-1344/Product: mucin I precursor, splice form A #status predicted <PREA>

F:1-62/Region: mucin I amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <SIGA>

F:1-19,29-33/Domain: signal sequence #link PREB #status predicted <SIGB>

F:1-19,29-1344/Product: mucin I precursor, splice form B #status predicted <PREB>

F:1-19,29-212,1033-1344/Product: mucin I precursor, epithelial tumor antigen splice form

F:138-1017/Region: 20-residue repeats (GSTAPPAHGVTSPDTRPAP)

F:1143-1344/Region: mucin I carboxyl-terminal non-repetitive

F:1245-1272/Domain: transmembrane #status predicted <TRM>

F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 9.2%; Score 672; DB 1; Length 1344;

Best Local Similarity 27.8%; Pred. No. 1.7e-21;

Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;

QY 238 LTVNKETTVEYKETTNNK--QTSDFGKEKTTSAKETOSIEKTSKAD-LAPTSKVLAKPT 294

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTTT-----KSAPTTTPKEPAPTTTKEP-----A 367

DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190

QY 368 PTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTT-----KKPAPTTTPKEP----- 413

DB 191 PDT--RPAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247

QY 414 --APTTTPKEPPTTTP-----KEPAPTTKEP-----APTTTPKEPAPTAPEK---- 450

DB 248 TSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGH 307

QY 451 KPAPTTTPKEPAPTTPK----EPAPTTTTPKEPAPTTT-----TKSAPTTTTPKEPAPTT 503

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTTT-----KSAPTTTPKEPAPTTTKEP-----A 367

DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190

QY 368 PTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTT-----KKPAPTTTPKEP----- 413

DB 191 PDT--RPAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247

QY 414 --APTTTPKEPPTTTP-----KEPAPTTKEP-----APTTTPKEPAPTAPEK---- 450

DB 248 TSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGH 307

QY 451 KPAPTTTPKEPAPTTPK----EPAPTTTTPKEPAPTTT-----TKSAPTTTTPKEPAPTT 503

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTTT-----KSAPTTTPKEPAPTTTKEP-----A 367

DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190

QY 368 PTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTT-----KKPAPTTTPKEP----- 413

DB 191 PDT--RPAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247

QY 414 --APTTTPKEPPTTTP-----KEPAPTTKEP-----APTTTPKEPAPTAPEK---- 450

DB 248 TSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGH 307

QY 451 KPAPTTTPKEPAPTTPK----EPAPTTTTPKEPAPTTT-----TKSAPTTTTPKEPAPTT 503

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTTT-----KSAPTTTPKEPAPTTTKEP-----A 367

DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190

QY 368 PTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTT-----KKPAPTTTPKEP----- 413

DB 191 PDT--RPAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247

QY 414 --APTTTPKEPPTTTP-----KEPAPTTKEP-----APTTTPKEPAPTAPEK---- 450

DB 248 TSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGH 307

QY 451 KPAPTTTPKEPAPTTPK----EPAPTTTTPKEPAPTTT-----TKSAPTTTTPKEPAPTT 503

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTTT-----KSAPTTTPKEPAPTTTKEP-----A 367

DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190

QY 368 PTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTT-----KKPAPTTTPKEP----- 413

DB 191 PDT--RPAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247

QY 414 --APTTTPKEPPTTTP-----KEPAPTTKEP-----APTTTPKEPAPTAPEK---- 450

DB 248 TSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGH 307

QY 451 KPAPTTTPKEPAPTTPK----EPAPTTTTPKEPAPTTT-----TKSAPTTTTPKEPAPTT 503

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTTT-----KSAPTTTPKEPAPTTTKEP-----A 367

DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190

QY 368 PTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTT-----KKPAPTTTPKEP----- 413

DB 191 PDT--RPAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247

QY 414 --APTTTPKEPPTTTP-----KEPAPTTKEP-----APTTTPKEPAPTAPEK---- 450

DB 248 TSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGHGVTSAPDTRRAPGSTAPPAGH 307

QY 451 KPAPTTTPKEPAPTTPK----EPAPTTTTPKEPAPTTT-----TKSAPTTTTPKEPAPTT 503

DB 15 LVLVLATTAPKPAIVTVTSGSHASSTPGGKEKTSATQRSVFSSTKNAVSMTSSVLSSHS 74

QY 295 P-KAETTTKGP--ALTTKPEP-----TPTTPKEPASTTP-----KE 327

DB 75 PGSGSSTTGQGDVTLAPATEPASGSAWQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134

QY 328 TPPTTT-----IKSAPTTTPKEPAPTT



Query Match	9.3%	Score 680;	DB 2;	Length 1229;
Best Local Similarity	26.6%	Pred. No. 7.3e-22;		
Best Overall Similarity	105.0%	Mismatches 394;	Indels 422;	Gaps 60;

Query Match	9.3%	Score 680;	DB 2;	Length 1229;
Best Local Similarity	26.6%	Pred. No. 7.3e-22;		
Best Overall Similarity	105.0%	Mismatches 394;	Indels 422;	Gaps 60;





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Db 1520 KASGTTASKVPVPAETQEVAVSSKRETFVAFVAFVAVSLCQMMRETFE

Matches 332; Conservative 114; Mismatches 515;

F.1350-1366/Domain: transmembrane #status predicted <TM2>

Matches 332; Conservative 114; Mismatches 515;

Db 1048 ASTEETVAPTEETVYABEETVPEETVYAPTEETVAPTEETVAPTEETVAPTEETVAPTE 1107  
Qy 783 TTPKETATTPKEAPATPKAPATTPPTPTTSEVSTPTTKEPT--TIHKSPDESTP 840  
Db 1108 ETVYAPAEETVPEETVYAPTEETVYAPTEETVYAPTEETVYAPTEETVYAPTEETVYAPTE 1166  
Qy 841 ELSAEPTPKA--LENSKPEPG-----VPTTKTPAAKPEMTTAKDKTTERDLRTTP--- 890  
Db 1167 YAPTEETVYAPTEETVPEETVYAPTEETVYAPTEETVYAPTEETVYAPTEETVYAPTEETV 1226  
Qy 891 ETTTAAAPKMKETATTEKTESKITATTTQVSTTTQDIT-----PKKIT--- 937  
Db 1227 EETVYAP-----TEETVPEETVYAPTEETVYAPTEETVYAPTEETVYAPTEETVYAPTEET 1281  
Qy 938 LKTTTLAPKVTTKTKTTITTEIMNKPEETAKPKDRAT--NSKATTPK--OKPTKAPKKPT 994  
Db 1282 TEETVYAPTEATVYAPTEETVYAPTEETVPEETVYAPTEETVYAPTEETVYAPTEETVYAPMEET 1341  
Qy 995 -----STKRP--KTMPRVRKPTTPTRKMTSTMPELNPTSRIAEAMLQTTTP 1041  
Db 1342 PYEPAEESTVSTVTEKPCNTEETDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPT 1399  
Qy 1042 --NQTSNKLVEVNPKSEDAG 1060  
Db 1400 CDQNGINGICVENKVRNNAG 1420

RESULT 4  
T16251  
hypothetical protein F35A5.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
C:Accession: T16251  
R:Leinbach, D.  
submitted to the EMBL Data Library, January 1996  
A:Description: The sequence of C. elegans cosmid F35A5.  
A:Reference number: Z18485  
A:Accession: T16251  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <LEI>  
A:Cross-references: EMBL:U46675; NID:g116613; PID:g1166621; PIDN:AA52641.1; GSPDB:GNOC  
A:Experimental source: strain Bristol N2; clone F35A5  
C:Genetics:  
A:Gene: CESP:F35A5.1  
A:Map position: X  
A:Introns: 1272/2

Query Match 11.6%; Score 846; Db 2; Length 1274;  
Best Local Similarity 27.6%; Pred. No. 7.2e-29;  
Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;  
Qy 76 PSSKKAPPSPGASQTIKTRSP-----KPP---NKKTKKVIIESEITIEHSVSENOES 128  
Db 185 PSPKKAAPSKEDHIVPTPIKPKAKKKWPPWEDDEVTEIEKEPEATRKVPAKKKEP 244  
Qy 129 SSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNK--KNRTKK--KP-----TP----- 176  
Db 245 STSVKVPVSPPTKKV-----PVKKEPEVPTPIKNTKTKKKWPPWEDDEVTEIEVK 294  
Qy 177 KPPVVDAGSGLONGD-----FKVTTPTDSTOHNVKSTSPKITTAKPI--NPRSLPPNSD 231  
Db 295 EPPVPEKAPVLLKDPAPAAKARDSPSKAAPKPKVEPSPVPPVPPVKNPKVKKYKPPWE 354  
Qy 232 TSKETSLTVNKEITVETKETTITNTKQSTDGKEKTSKETSIAKTSADLAPTSKVL 291  
Db 355 VDDEPAEAVKPKPAEKKTPVLKRKEPEPSSSTPPSDSPKKAAPAVKPRDSPSKKATPL 414  
Qy 292 KPTPKAE-----TTTKGPA-----LTTTPKEPTTP-----KEPATTPK 326  
Db 415 QADPKAQEPPTPVKNPKVKKYKPPWEDDEVTEIEVKQPEAPKATPVLLKRKEPAKDTA 474

Qy 327 EP-----TP--TTTKSAPTTPKEAPATTTKSAPTTPKEAPATTTKEPA----- 367  
Db 475 KPATSKTPTPEKPKDPVKPRDSSPKKVAAPKDSQAQAP--TPVKNPVKKWRPWEDEETPA 533  
Qy 368 -----PTTPEAPATTTKEAPAT-----TTKSAPTTPKE-----APTTPKKAPATTP 410  
Db 534 DVSKSPTDAKTPSLAKKDPAPAKESLKPADTKAPKPRDPSPKKVPAPAKPTAPKTPVLA 593  
Qy 411 KE-----PAPTTPKEPTTPKEAPATTTKEAPATTPK--EPAPTAPKKAP-----TTPKE 459  
Db 594 KKEPAGADSKTKEKESKPRDPSPKKAVPAKVPKTEVAPAAVKKPEPISKPKDTPAKK 653  
Qy 460 PAPTTPKEAPATTTKEP-----SPTTPKEAPATTTKSAPTTPKEAPATTTKSAPTTP--KEP 514  
Db 654 AEPNSPVVP--PTPVKNPVKKWKPWEDDDAPAKPVSLPEPEKK--TPVLAKKAPTTPDSEA 711  
Qy 515 SPTTPKEAPATTPK--EPAPTTPKKAP--TTTPKEAPATTTKEAPATTTKEAPATTTKEAPT 678  
Db 712 AADPVSGSPSSDKPLAKKAPVKPRDPSMKAVPIKPAKT--EYPPAVVKKPEPVAKSRD 769  
Qy 570 PAPTTPKETATTPPKLTPTTP-----EKLAPTTPKEAPATTPPELAPTTPPEP-----T 619  
Db 770 PSPKAK--AEPNSP--VVPPTPVKNPVKKWKPWEDDDAPAEPVNVPEPEKKTPVLAKKT 826  
Qy 620 PTTPEEAPATTPKAAAPNTPEAPATTPKEAPATTTKEAPATTTKEAPATTTKEAPTTPKT 678  
Db 827 PVKPRDPSPKKAVPAKPTKTDAPPVSVKPEPVSKPEPSPKKAEPSNVVP-----PTP 882  
Qy 679 LKEAPATTPKPAKELAPT--TTKEPTSTSDKPAATTPKGTATTPKPEAPATTPKPEAP 737  
Db 883 VKNPVKKW--KPPWEDDDDETEEVKPSSE--PEKKTTPVLAK--KEPEKPKD--APKVAAPK 937  
Qy 738 TTPKGTATPTTLKEAPAT-----TPKKAP-----KELAPTTPKPTS-----TTSK 779  
Db 938 PSPKAVPE--KEPAKVAAPKPRDLSPKKAIPIANTQEAAPTTPVKNPVKKWKPWEDDD 995  
Qy 780 P-----APTTPKET-----APTTPKEAPATTPKPAATTPPTTPPTTSEVSTPTTKEP 828  
Db 996 PAEPVSAPEPEKTPVLAKKAPAKPRDP---SPKKAAPVAAK--PDKPIEV--PPTPVKNP 1050  
Qy 829 TTIHSP-----DESTPELSA--EP---TPKALENSKPEGV-----PPTTKPA 867  
Db 1051 VKKWKPPWEDDDDEPSPVSAPEPEKTPVLAKKAPTATKPKDSEAAADPVSGPTSKDPK 1110  
Qy 868 AT-----KPEMTTAKDKTTERDLRTTPETT--TAAPKMTKETATTTTEKTTESKITATT 919  
Db 1111 LSKKAPVEKPKPTDPKDDKLPSPAKKPEKAPAEAPKPKWPPWDDDPDEPEADFTVPA 1170  
Qy 920 TQVSTTTQDTPPFKITTLKTTTLAPKVTTTKTTITTEIMNKPEETAKPKDRATNSKAT 979  
Db 1171 PSKKPDTEDPADPLG-----GPKTKDPK-----LNKKAPEKPTK----- 1206  
Qy 980 TPKPOKPTKAPKKPTSTKPKTMPRVKPK-----TTPTPKMTSTWP 1022  
Db 1207 -PKPEVSKPEPKPEPEPKP--AAPKKWPPWEDDDPDEPEADFTMPAPKKPDTEDP 1259

RESULT 5  
T30826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N:Alternate names: alpha-NAC protein  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999  
C:Accession: T30826  
R:Yotov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1763-1772, 1996  
A:Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a mus-  
A:Reference number: Z20889; MUID:96312450  
A:Accession: T30826  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2187 <YOT>



J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.  
A:Reference number: A61257; MUID:91086481  
A:Accession: A61257  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>  
A:Experimental source: bronchus  
R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner, Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus  
A:Reference number: P00328; MUID:92198477  
A:Accession: P00328  
A:Molecule type: mRNA  
A:Residues: 2328-2468 <XUG>  
A:Cross-references: GB:M86523  
A:Experimental source: small intestine  
A:Accession: P00329  
A:Molecule type: protein  
A:Residues: 2328-2342, K',2344-2354 <XUG1>  
C:Genetics:  
A:Gene: GDB:MUC2  
A:Cross-references: GDB:120203; OMIM:158370  
A:Map position: 11p15.5-11p15.5  
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology  
C:Keywords: glycoprotein; intestine; tandem repeat  
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>

Query Match 14.7%; Score 1073; DB 2; Length 3020;  
Best Local Similarity 27.1%; Pred No. 4.3e-38;  
Matches 374; Conservative 101; Mismatches 502; Indels 402; Gaps 50;

QY 29 SCAGRGEGYSDATNCNDYCNQHWECD-----CPDFKRVCTAAEHNPTS 74  
Db 1079 SCS--CDTG--GDCEFCGSAVYAQECTKEGACVFWRTPDLCPIF-----CDYINPPH 1128

QY 75 PPSKKAPPPGASQTIKSTT-----KSPK--PNKKKKTKVIESEB-- 115  
Db 1129 ECEWHYEPGCGNSFETCRINGIHNSISVLEGYPCPCDRIYEDLKKCVTADKCG 1188

QY 116 ---ITEEH---SVSENOQSSSSSSSSSIWIKSKNSAANRELQKLVKDNKNR 169  
Db 1189 CVVETHPPGASVPTTECTSCVCTNSQVVCREEGK-----ILNQ 1231

QY 170 TKKKTPKPPVVVDAGSLDNGDF---KVTPDTSSTQH---NKVSPKITTAKPNR 223  
Db 1232 TQ-----DGAFCYWEICGPNGTVEKHFNICITTRPSLTFTTTITL 1273

QY 224 PSLPNSDTSKETSIVNKKETVETKET--TTTNK-----QTSTDGKEK 266  
Db 1274 FTTP-----TSFTTTTTTTTTTSSTVLSLTPKLCCLSDWINEHPSSGSDGDRP 1325

QY 267 TS--AKETQSIKTSKAD----- 282  
Db 1326 FDGCGAPEDIECRSVKDPHLSLEHQCKVQCDVSVGFIKCNEDQFGNGPFGICYKIR 1385

QY 283 -----LAPTSKVLAKPTPKAETTTKGPALTTPKEPTTPPKPEASTTPKEPTPT 332  
Db 1386 VNCCWPMKCIITPSPTTTPSPPTTTTLPTTTPSPPTTTTTPPTTTPSPPTTTT 1445

QY 333 IKSAP--TTPKEPAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAP 391  
Db 1446 TTPLTTPSPPTSTTTTPPTTPSP--PTTTPSPPTTTPSPPT--TTTTPPTTTPSP 1503

QY 392 --TTPKEPAPTTKPPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 450  
Db 1504 MTPPTTPPAATTLPPPTTTPSPPTTTTTPPTTTPSPPTTTPITP--PSTTTTLPPT 1562

QY 451 KPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAAP--TTPKEPAPTTKSA 508  
Db 1563 SPPPTTTPPTTTPSPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTPSP 1622

QY 509 TTPKEPSPTTTPKEPAPTTTPKEPAPTTTPKAPPTTPKAPPTTPKAPPTTTKKKAPTAPK 568  
Db 1623 TTPITPPTSTTTTPPTTTPSPPTTTTTPPTTTPSPPTTTTTPPTTTPPTTTPPTTTP 1682

QY 569 EPAPTTTPKETAPTTPKKTTPTEKLAPTTPKAPPTTPPEELAPTTPEEPTTPPTPEEPAP 628  
Db 1683 SPITTP--SPPTT---TWTPS--PTTTPSSPTTTPPTTTPSSSTTP--SPPTTMTTTPSP 1733

QY 629 TTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTP--RETAPTTTP-- 671  
Db 1734 TTPSPPTTMTLPTTTPPTTSSPLTTLPLPSPITPPTPESPTSTTPPTPCVPLCWNWGLD 1793

QY 672 KG----- 673  
Db 1794 SGKPNFHKPGGDTGLDVGCGFWAANISCRATWYDVPVIGQLGQTVVCDVSVGLICKNE 1853

QY 674 -----TAPTTLKAPPTTPPKKAPKELAPT--TKETS 705  
Db 1854 DOKPGVIVPMAFCLNYEINVQCCECVTQPTM-----TTTTTENPTPTTITTTTTPPTP 1910

QY 706 TTSOKP-----APT--TPKGTAPTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKAPPTTPKKPA 760  
Db 1911 TSTQSPNGLQAPTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1969

QY 761 PKELAPTTTKGPTST--TSDKAPPTTPKETAPTTPKEPAPTTTPKKP-----APTTPETP 812  
Db 1970 TVLITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2029

QY 813 PPTTSEVSTPTTK-----EPTTIHKSDESTPELSAETTPKALENSPKPGVPTTKTPA 867  
Db 2030 TPTGTSGTPHTSTADIACLTSSNPPPESTPTQTSRSTSSPLTESTTLLSTLPAIEMT 2089

QY 868 ATPEMTTAKDKTERDLTTPETTAAPKMTKETATTEKTTESKITATTTQVNTSTT 927  
Db 2090 STAPPTTAPTTPSGHTLSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2147

QY 928 QDTPPKITT---LKTTLAPKVTTTKITTEINKPEE----- 965  
Db 2148 WTPPTPLSPSILRTGLRPSPSVLICVLDNTYVAFGEVYNGYDTCYFVNCSL 2207

QY 966 -----TAKPKDRATNSKATTPKPKQTKAPKKTSTTKPKTTPRVRPKPTTTPR 1015  
Db 2208 CTLEFYNWSCTSPSTPTTPSK--STPTSPKSPSTSPKPTPGTKPCEPDPDFPR----- 2260

QY 1016 KMTSTPELNPTSRIAEAMLOTTTRPNQTPNSKLVEVNP-----KSDAGG- 1061  
Db 2261 -----QENETWLCDCFN--ATCYNTNTVEIVKVECEPPMPPTCSNGLQPVVRVEDPDG 2312

QY 1062 -----AEGETPHMLL-----RPHVFPEVTPMDYLPVPRVNOGIIN 1098  
Db 2313 CWHWECDYCTGWGD--PHYVTFDGLVYSVQGNCTVVLVEISPSVD-----NFGVYID 2364

RESULT 2  
T18262  
S-layer protein - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18262  
R:Fujino, T.; Beguin, P.; Aubert, J. P.  
J. Bacteriol. 175, 1891-1899, 1993  
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose  
A:Reference number: 218847; MUID:93209931  
A:Accession: T18262  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1664 <UJ>  
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 13.0%; Score 950; DB 2; Length 1664;  
Best Local Similarity 31.0%; Pred. No. 3.8e-33;

Mon Apr 29 08:35:21 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:31:52 ; Search time 114.61 Seconds  
(without alignments)  
907.900 Million cell updates/sec

Title: AA6  
Perfect score: 7294  
Sequence: 1 MANKTLPYLLLLLSVFVQ.....ARATTRSGTSLKVMYNCP 1366

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1073	14.7	3020	2 A43932	musin 2 precursor,
2	950	13.0	1664	2 T18262	S-layer protein -
3	863.5	11.8	1489	2 T31108	cyst germination s
4	846	11.6	1274	2 T16251	hypothetical prote
5	812	11.1	2187	2 T30826	nascent polypeptid
6	810.5	11.1	1367	1 S48478	glucan 1,4-alpha-g
7	784	10.7	1188	2 S49915	extensin-like prot
8	695	9.5	1151	2 T18535	high molecular mas
9	680	9.3	1229	2 T25697	hypothetical prote
10	672	9.2	1344	1 A35175	musin 1 precursor,
11	661.5	9.1	7962	2 T38346	elastic titin - hu
12	649	8.9	3570	2 T45025	musin MUC5B, trach
13	647.5	8.9	3507	2 T34513	hypothetical prote
14	643.5	8.8	990	2 T51618	nucleolar phosphop
15	633	8.7	489	2 T11622	extensin class 1 p
16	632	8.7	761	2 C84672	hypothetical prote
17	629	8.6	6642	2 T23757	protein UNC-89 - C
18	627.5	8.6	971	2 T19431	hypothetical prote
19	622.5	8.5	839	2 F75518	hypothetical prote
20	607.5	8.3	801	2 T29018	hypothetical prote
21	607.5	8.3	924	2 S27923	gene LF3 protein -
22	605	8.3	379	2 S50125	larval glue protei
23	592.5	8.1	2232	2 T34434	hypothetical prote
24	568.5	7.8	1630	2 A53577	ascites stialoglyco
25	559.5	7.7	350	2 S22456	hydroxyproline-ric
26	559	7.7	875	2 S23760	polyphenolic adhes
27	556.5	7.6	856	2 T16543	hypothetical prote
28	551	7.6	620	2 S06733	hydroxyproline-ric
29	551	7.6	1087	1 QFMSH	neurofilament trip

30	545	7.5	873	2 A47283	calphotin - fruit
31	542	7.4	369	2 S20500	hydroxyproline-ric
32	538.5	7.4	416	2 JU0465	extensin precursor
33	537.5	7.4	1459	2 T32271	hypothetical prote
34	537	7.4	1072	1 A37221	neurofilament trip
35	532	7.3	756	2 T27642	hypothetical prote
36	530.5	7.3	813	2 S70795	vsaa protein precu
37	530.5	7.3	1162	2 JH0557	exo-alpha-sialidas
38	530	7.3	865	2 A47282	calcium-binding pr
39	526	7.2	3534	2 T42567	tegment protein 2
40	522	7.2	328	2 JQ0985	hydroxyproline-ric
41	521.5	7.1	854	2 S02003	neurofilament trip
42	518	7.1	866	2 T45462	membrane glycoprot
43	518	7.1	1611	2 T38236	hypothetical prote
44	513	7.0	1832	2 T31113	musin-like glycopr
45	512.5	7.0	867	2 T45463	membrane glycoprot

ALIGNMENTS

RESULT 1  
A43932  
musin 2 precursor, intestinal - human (fragments)  
N:Alternate names: musin SMUC-41  
C:Species: Homo sapiens (man)  
C:Date: 10-Mar-1993 #sequence\_revision 12-Apr-1996 #text\_change 05-Nov-1999  
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A49963; MUID:94132002  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GU1>  
A:Cross-references: GB:L21998  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MUID:93016075  
A:Accession: A45106  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398  
A:Experimental source: colon  
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,  
J. Clin. Invest. 88, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MUID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi  
A:Reference number: A33532; MUID:89197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874  
A:Experimental source: intestine  
R:Janjy, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

```

QY 837 TTTKEPTT-----IHKSPDESTPELSAEPKPALENSKPEGCVTTTKTPAA----- 882
Db 703 DXRXPXPGSTAPXAHGVTSA-PDXRXPXPGSTAPXAHGVTSA-PDXRXPXPGSTAPXAHGVTSA-P 762
QY 883 -TRPEWTTAKDK---TTERDLRTTPTTTAAPKMTKETATTTTEKTESKITATTTQVTS 938
Db 763 DXRXPXPGSTAPXAHGVTSA-PDXRXPXPGST--APXAHGVTSA-PDXRXPXPGSTAPXAHGVTSA 820
QY 939 TTTQDTTPFKITTLKTTTLAPKVTTTKTTTTEIMNKPEETAAPKDRATNSKATTPKQP 998
Db 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTSA-PDXRXPX- 868
QY 999 KPTKAPKPTSTKKPTMPRVRRKPTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNOT 1058
Db 869 -GSTAPXAHGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA 911
QY 1059 -PNSKLEVNPKSEDAGAEGETPHMLLRPHVFMPEVTPDMDYLPRVPNQGIINPMLSD 1117
Db 912 APXAHGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA 957
QY 1118 ETNICNGKVPDGLTT-LRNGTLVAFRGHYFWMMLSPFSPARRITEVWGIPSPID 1172
Db 958 VTSAPDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA-PDXRXPXGVTSA 1003

```

Search completed: April 26, 2002, 16:16:52  
Job time: 203 sec

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE: Peptide  
NAME/KEY: 144  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,  
OTHER INFORMATION: or AGC; and Asn = AAT or AAC."

FEATURE: Peptide  
NAME/KEY: 147  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE: Peptide  
NAME/KEY: 1..21  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-2

Query Match 5.7%; Score 424.5; DB 4; Length 2035;  
Best Local Similarity 23.7%; Pred. No. 2.4e-18;  
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 250 TSLAVNKETTVETKETTNNKQSTDCGKETSATQSIKTSKAD-LAPTSKVLAKPT 308  
Db 16 TVLTV-----VTGSHASSTPGGKETSATQSSVPSSTEKNVAMTSSVLSHS 65  
QY 309 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTPKEPTTTIKSAP 351  
Db 66 PGSSSTTGGQDVTLAPATEPASGAATWGQDVTSVPTRPALGSTTPPAHDVT---SAP 122  
QY 352 TTPKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 392  
Db 123 --DNKPAFGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 180  
QY 393 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPKEPTPTT 439  
Db 181 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 240  
QY 440 PKE--PAPTTTPKEP-----APTTPKEPAPTAPK-----KPAPTTTPKEPAPTTPKEPAPT 486  
Db 241 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 300  
QY 487 TKEPSTTPKEPAPT--TKSAPTTTPKEPAPT-----TKSAPTTTPKEPAPTTPK-----E 535  
Db 301 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 359  
QY 536 PAPTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPK-----EPAPTTTPKPAKPAPTAP 587  
Db 360 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 419  
QY 588 TPKEPAP-----TTPKKLTPTTPEKLPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 643  
Db 420 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 471  
QY 644 TPKA-----AAPNTPKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKPGT--APTTLKEPAP 598  
Db 472 APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGS 530  
QY 699 TTP-----KKPAKELAPTTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 746  
Db 531 TAPXAHGVTSPADXPXPGSTAP--XAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGS 595  
QY 747 KEPAPTTPKG-----TAPTTLKEPAPTTP-----KKPAKELAPTTTTPKEPAPTTP 792  
Db 586 XPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAP--XAHGVTSPAD 644  
QY 793 KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTTPETPTTSEVSTP 836  
Db 645 RPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGS--TAPXAHGVTSP 702

Db 912 APXAHGVTSPADXPXPGSTAPXAH-----GVTSAPDXRXPXPGSTA-----PXAHG 957

QY 1118 ETNICKGKPVGLTT--LRNGTLVAFRGHYFWMLSPFPSPARRITEVWGIPSPID 1172  
Db 958 VTSPADXPXPGSTAPXAHGVTSA-----PDXRXPXPGSTAPXAHGVTSPAD 1003

RESULT 15

US-09-083-116-2  
Sequence 2, Application US/09083116  
Patent No. 6203795  
GENERAL INFORMATION:  
APPLICANT: CHAMON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LAHE, Richard  
APPLICANT: HAREUVENI, Mara  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,116  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/479,537  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1899  
OTHER INFORMATION: /note= "The amino acids spanning  
128 to 1899 constitute a repeated region wherein the repeat  
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,



NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,537A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 90/13101  
FILING DATE: 23-OCT-1990  
APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1899  
OTHER INFORMATION: /note= "the amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."  
OTHER INFORMATION: repeats varies from 1 to 40."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA, or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 5.7%; Score 424.5; DB 2; Length 2035;  
Best Local Similarity 23.7%; Pred. No. 2.4e-18;  
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;  
QY 250 TSLTVNKETTVEKTTTNNKQTSDGKEKTTSAKETQSTIEKTSAD-LAPTSTKVIAKPT 308  
DB 16 TVLTV-----VIGSCHASSTPGGKETTSATORSVSPSTTEKNAVSMTSVLSSH 65  
QY 309 P-KAETTTKGP--ALTPKPEP-----TPTPKPEASTTPKEPTTTIKSAP 351  
DB 66 PGSGSSTQGGDVTLPAPATEPASGSAATWQDVTSPVTRPALGSTTPAHVDV---SAP 122  
QY 352 TTPKEPAPTTT-----KSAPTTKPEP-----APTTTKEPAPTTPKEPAPT 392  
DB 123 --DNKPAPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 180  
QY 393 TKE--PAPTTT-----KSAPTTKPEPAPTTPK-----KPAPTTKPEAPTTPKEPTPT 439  
DB 181 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 240  
QY 440 PKE--PAPTTKPEP-----APTTTKEPAPTAPK-----KPAPTTKPEAPTTPKEPAPT 486  
DB 241 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 300  
QY 487 TKEPSPTTPKEPAPT--TKSAPTTTKEPAPT-----TKSAPTTKPEPSPTTK---E 535  
DB 301 APDXRP-XPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 359  
QY 536 PAPTTPKEPAPTTPK-----KPAPTTKPEPAPTTPK-----EPAPTTTTPKAPTAPKEAPT 587  
DB 360 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 419  
QY 588 TPKETAP-----TPPKLTTTPEKLAPTTPEKAPAPTPEELAPTTPEEPTTTTPEEAPT 643  
DB 420 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 471  
QY 644 TPKA-----AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGT-APTTLKEPAP 698  
DB 472 APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 530  
QY 699 TTP-----KKPAPKELAPTTTKEPTSTSDKPAP---TTPKGTAPTTPKEPAPTTP 746  
DB 531 TAPXAHGVTSPADXPXPGSTAP-XAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 585  
QY 747 KEPAPTTPK-----TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPTSTSD 792  
DB 586 PXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 644  
QY 793 KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTPETPTTSEVSTP 836  
DB 645 RPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 702  
QY 837 TTTKEPTT-----IHKSPDESTPELSAPETPKALENSPKPCGVPTTKTPAA----- 802  
DB 703 DXRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 762  
QY 883 -TKPEMTTAKDK---TTERDLRTTPTTTAAPKMKTKETATTTETTESKITATTTQVTS 938  
DB 763 DXRPXPGSTAPXAHGVTSPADXPXPGST--APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 820  
QY 939 TTTQDTTPTTKITTLTKTTLPAPKVTTKKTLTTTTEIMNKPEETAKPKDRATNSKATPKPKQ 998  
DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 868  
QY 999 KPTKAPKPTSTKKPTMPRVKPKTTPTPRKMTSTMPNLNPTSRIAEAMLQTTTTPNOT 1058  
DB 869 -GSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 911  
QY 1059 -PNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPMDVLPVPRVNOGIINPMLSD 1117

APPLICATION NUMBER: 08/479,537  
FILING DATE: 23-OCT-1991  
PRIORITY DATE: 23-OCT-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-6021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1727  
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat of 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."  
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."  
US-09-083-116-5  
Query Match 5.7%; Score 424.5; DB 4; Length 1867;  
Best Local Similarity 23.7%; Pred. No. 2.1e-18;  
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;  
QY 250 TSLVTNKETTVEKTTTNTKSTGKETTSAKETQSIKTSKAD-LAPTSKVLAKPT 308  
DB 16 TVLTV-----VTGSGHASSTPGGEKETSATQSSVPSSTKNAVMTSSVLSSHS 65  
QY 309 P-KAETTTKGP--ALTTPKEP-----TPPTPKPEASTTPKEPTTIKAP 351  
DB 66 PGSGSSTQGDVTLAPATAPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVT---SAP 122  
QY 352 TTPKEPAPTT-----KSAPTTPKEP-----APTTPKEPAPTTTPKEPAPT 392  
DB 123 --DNKPAPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 180

393 TKE--PAPTTT-----KSAPTTTPKEPAPTTTPK-----KPAPTTTPKEPAPTTTPKEPTTT 439  
DB 181 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 240  
QY 440 PKE--PAPTTTKEP-----APTTPKEPAPTPAK-----KPAPTTTPKEPAPTTTPKEPAPT 486  
DB 241 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 300  
QY 487 TKEPSTTPKEPAPTT--TKSAPTTTTPKEPAPTT-----TKSAPTTTTPKEPSTTTTK-----E 535  
DB 301 APDXRP--XPGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 359  
QY 536 PAPTTTPKEPAPTTTPK-----KPAPTTTPKEPAPTTTPK-----EPAPTTTTPKEPAPTTTPKEPAPT 587  
DB 360 APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 419  
QY 588 TPKEPAP-----TTPKLTTPTEKLAAPTTTPKEPAPTTPEELAPTTTPPEPTTPPEEPAPT 643  
DB 420 SAPDXRXPXGSTAPXAHGVT-----APDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 471  
QY 644 TPKA-----AAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGT--APTTLKEPAP 698  
DB 472 APXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 530  
QY 699 TTP-----KKPAPKELAPTTTPKEPAPTTTPKEPAPTTTPKGT--APTTLKEPAP 746  
DB 531 TAPXAHGVTAPDXRXPXGSTAP--XAHGVTAPDXRXPXGSTAPXAHGVT-----APDXR 585  
QY 747 KEAPATTTPKG-----TAPTTTPKEPAPTTTP-----KKPAPKELAPTTTPKGTSTTSD 792  
DB 586 XPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 644  
QY 793 KPAP--TTPK-----ETAPTTTPKEPAPTTTP-----KKPAPTTTPPTTPPTTPSEVSTP 836  
DB 645 RXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 702  
QY 837 TTPKPTT-----IHKSPDESTELSAEPTPKALENSPKPEPGVPTTKTAA----- 882  
DB 703 DXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 762  
QY 883 -TKPEMTTAKDK---TTERDLRTTTPETTAAPKMTKTATTTEKTESKITATTQTTS 938  
DB 763 DXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 820  
QY 939 TTTQDTPPEKITTLKTTLLAPKVTTKTTTTEIMNPEETAETPKDRATNSKATTPK 998  
DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 868  
QY 999 KPTKAPKPTSTKPKTKMTPRVRKPKTTTPRKMSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMT 1058  
DB 869 -GSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTSA-PDXRXP-----PGST 911  
QY 1059 -PNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVPRVNOGIIINPMLSD 1117  
DB 912 APXAHGVTAPDXRXPXGSTAPXAH-----GVTAPDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 957  
QY 1118 ETNICKPKVDGLTT-LRNGTLVAERGHYFWMLSPSPSPARRITEVWGIIFSPID 1172  
DB 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTAPDXRXPXGSTAPXAHGVT 1003

RESULT 14  
US-08-479-537A-2  
; Sequence 2, Application US/08479537A  
; Patent No. 5861381  
; GENERAL INFORMATION:  
; APPLICANT: CHAMON, Pierre  
; APPLICANT: KIENY, Marie-Paule  
; APPLICANT: LATHE, Richard  
; APPLICANT: HAREUVENI, Mara  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR



Mon Apr 29 08:35:42 2002

```

; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-041-886-23

Query Match          5.7%; Score 425.5; DB 4; Length 1185;
Best Local Similarity 21.8%; Pred. No. 1 le-18;
Matches 240; Conservative 119; Mismatches 437; Indels 303; Gaps 52;

QY 102 SOTIKSTTKRPPKPKKTKKVKIESEITEHSVSNQESSSSSSSSSSSIWIKSS 161
Db 9 SMSRSGRKKKAPGR-----BELSRGRASPGVSTSSDG-----KAES 50
QY 162 KNSAANRELQKLVKDKNKKTKKTPKPPVVDVDEAGSLDNGDFKVTPTDSTTOH 221
Db 51 ROTA-----KKARVEEASTPK---VNKGRS-----EISESESE 82
QY 222 VSTSPKITTAKINPRPLPNSDTSKETSLTVNKTETVETKTTNTKOTSTDGKETT 281
Db 83 ETNAPKTKTEGELPPRPSDLSLGRSL-NDGSSDRDIDQDNKST-----P 133
QY 282 SAKETQSIEKTSAKDLATPKTAKTTPKATTKGPAITTKPEPTTP--KEPASTTP 339
Db 134 SIYSPGSEVDS-----DSSGLSOGPA--RYPHPPLFPSPQPPDSTP 176
QY 340 KEPTPTTKSAPTTKEPAPTTTKSAPTTKEPAPTTTKPAPTTTKPAPTTKEPAPT 399
Db 177 ROPEASF-----EPHPSVTPTGYHAPMEP-PTSRMFQAP--PGAPPHPQLYPGGT 224
QY 400 --TTKSAPTTPK-----EPATTPKKAPTTTPKEPAPTTTPKEPTTPPK 441
Db 225 GGVLSGPPMGPKGGGAASVGGPGNGKGQHPPTTPLISVSSSGASGAPTKP----PTTPV 280
QY 442 EPATTTKEAP-----TTKEPAPTA--PKKPAPTTKEPAPTTKEPAPTTKEPAPT 494
Db 281 GGNLPSAPPANFPHVTNLPPLPPALPLNNASASP-----PGLGAQPLPGHL--PSPYA 334
QY 495 PKE-----PAPTTTKSAPTTKEP--APTTKSAPTTT--KEPSPTTTKEPAPTTKEPA 545
Db 335 MGQGMGLPGP---EKGPPLAPSHPLPASSAPAPPMRFPYSSSSSSSSSSSSSS 391
QY 546 PTPPKKAPTTKEPAPTTKEP--PAPTTTKKAPTAAPKAPPTTPKETAAPTTPKKIT 601
Db 392 SSSASPPFAS--QALDSYHSPPTSL-----SVSNQP---PKYQPSLPQAVWSQ 440
QY 602 --PTTPE--KLAPTTKEPAPTTPEELAPTTPEEPTT----- 635
Db 441 GPPIPPYGRLLANSNAHPGFPFSTGAQSTAHPPVSTHHHHHQQOQQOQQOQQHHG 500
QY 636 --TPEEAP-----TTPKAAPN-----TPKEPAPTTTPKEP-----APTTPKE 671
Db 501 NSGPPPPGAFPHPLEGGSSHHADYAMSPSLGRPLPPYPPGPAHLPPPHSQVSYSQAGP 560
QY 672 PAPTPKETAPTTPKGTAPTTKEPAPTTKPKAPKELAP-----TTTKEPTSTSDKAPT 728
Db 561 PVSSSSSSSSTSGSYPCS--HPSPSQGQAGAPYPPVPTVTTSSATLSVIATVAS 618
QY 729 TPKGTAPTTKEPAPTTTPKEPAPTTTPKGTATTLKEPAPTTTPKAPKELAPTTTKGPTS 788
Db 619 SPAGYKTASPPGPPYGRKRAPSPGAYKATPPGKPGSP-----PSPETGTPPGYRG--- 670

```

```

RESULT 12
US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: KIENY, Marie-paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835
; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021

```

QY	454	TPKEPAPAPAKKPA--PPTPKERA--PPTTKEPARTTTKEPSPTP-----KEPAPTTTK	504
Db	197	YP-----PTTKAKPSYPPPTTKAKPSYPPPTTKAKPSYPSLTKAKPSYPTTK	252
QY	505	SAPTTTKERA-----PPTTKSAPTTKPEPSPTTKKEPA--PPTPKERA-----PPTPKK	551
Db	253	AKPIYAKPTTKAKPTVPTTKAKPTVP-----PTYKAKPSYPPPTTKAKPTTKAKPTTKAK	308
QY	552	PA--PPTTKERA--PPTTKPEPAPTTTKPAPTAAPKPEPAPTPKETAPPTPKLTPPTPEK	607
Db	309	PSYPTTKAKPSYPPPTTKAKPSYPPPTTKAKPSY-----PTYKAKPSYPPTKVKPT--YK	362
QY	608	LAPTTPE--KPAPTTPEELAPTTPEBPT--PPTPEERA--PPTTKAANPTKPEAPTTPK	662
Db	363	AKPIVPTSYKAKPSY-----PTTKAKPSYPPPTTKAKPSYPPPTTK--AKPSY-----PTYKA	413
QY	663	EPA--PPT-----PKEPAPTTPKETAPTT--PKGTAPTTLKEPAPTPPKPAPKELAPTTTK	715
Db	414	KPTVPSYTKAKPSYPPPTTKPKISYPTTKAKPSYPSYTKAKSSYPTTKAKPSYPTTKA	473
QY	716	EPTSTTSKCAPTPTKGTAPTTPKPEPAPTTKPEAPTTPKGTAPTTLKEPA--PPTPKKP	773
Db	474	KPTVPSYKAKPTYK--AKPSY-----PTYKAKPS-----YPTTKAKPSYPTTKAKP	521
QY	774	APELAPTTTKGTSTTSOKPAPTTPKETAPTTKEPA--PPTPKKPA--PPT-----PPTP	826
Db	522	TYK--AKPTYKAKPSYPPPTTKAKPSY-----PTYKAKPSYPPPTTKAKPTVPTTKAKPSY	576
QY	827	PPTTSYSTPTTKKEPTTLHKSDTESTPBLSEPTPKALENSKEPBGVPTTKTPAATKPE	886
Db	577	PTYKAKPSYPT-----YKAKPSYPPPTTKAKPTTKA-----KPTVPT-----YKAKPS	620
QY	887	MTTAKDKTTERDLRTPPTTTAAP	911
Db	621	YPPTHKAKPTTKAKPSYPPPTTKAKP	645

RESULT 10

```

US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUTSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219, 849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-5

```

Query Match 6.1%; Score 452; DB 4; Length 960;  
Best Local Similarity 25.6%; Pred. No. 1.9e-20;  
Matches 173; Conservative 60; Mismatches 309; Indels 134; Gaps 41;

[illegible]

; APPLICATION NUMBER: US/09/041,886



Qy	376	TTKEPAPTTTKEPAPTTTKEPAPTTTTSAPTTTKEPAPTTTPKKPAPTTTKEPAPTTTKEPAPTTTKEP	435
Db	272	TT	331
Qy	436	TTPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT	495
Db	332	TTTTTTKKPTT	377
Qy	496	KEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTT	527
Db	378	TKPTTT	437
Qy	528	-----PSPTT-----	535
Db	438	VICKDRIENGMAFTWIPNDTHVFRFKVDVNTISVRCRGAGKLEFPDPSLDTFTT	497
Qy	536	P-----APTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	559
Db	498	PVAGHNSCSIIIVSGSDGKIHVSPVSGSDVSLIS-----AFIQSELFNEVYCDCTAKYG	553
Qy	560	---PAPTTEKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	616
Db	554	AHSGVQTSADVTTTTTAKPTTTT--TGAPCQPTTTTGTGSPSKPTTTTTTTTAKTTT---	607
Qy	617	APTTEELAPTTTPEEPTPTTPEEPAPTTTPKAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	676
Db	608	--TTLNPIITTTTQKPTTTT-----TT--KVGPKPIATTTTTLKPIVTTTTTTTAKTTT	656
Qy	677	PRETAPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	719
Db	657	TTTTTPTT-----TTTTKRDEMTTTTTPLDIDGIEITPIPIEKMLDKYRMIYDYNSSL	711
Qy	720	--TTSDDKAPPTTPKG--TAPTTEKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE	773
Db	712	LLDSNDEPIGSGAQIADTSLNLFVQTHKSTGLPIDPMVGLFDPDKSGNLVHPYNTQM	771
Qy	774	APKELAPTTTKGPTTSTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	833
Db	772	SGLSVSYLAANKLVDDDEYTG--LPIDTILGTGLDPVSLIPFN--PETGELFOPISDEI	827
Qy	834	SPTT-----TTKEPTTTHKS-----	853
Db	828	MNGTIAGIVSGISASELSLQKALIDPATNMVWVGEFGLLNATGVMPIGFLGPSBQQT	887
Qy	854	-----PELSAETPKALENSKPEGPVPTTKTPAATKPDMTTAKDKTTER---	898
Db	888	FSPEIEDGGIIPPEVAANADKFLSIP--PSVP-----ESIEP-----KDKIDSISE	934
Qy	899	-----DLRTPETTTAAPKMTKEATTTEKTTESKITATT--TQVITSTTQD	943
Db	935	LWYDIESGRLLIGQVSKRPIPGSIAGDLNIPMKPTTQTDSTVTKPIDTPTGLFPNPTGHL	994
Qy	944	TTPKKITLTKTTLAPKVTTTKTITTTTETIMNKP--EDTAKPKD-----RATNSKATT	994
Db	995	INPTNNMTDSSFAGAYKVAVSNIGIKTDNVYGLPVGELTGPKDGPDSIDFNSSTGELVD	1054
Qy	995	PKPKPTKPKKPTSTTKPKTMP-----	1033
Db	1055	PSGCKPILNNSTAGIVSGKPLGPPIEDENGLNFDPSNLPIPDGNLNQVNPETNSTVSGSTS	1114
Qy	1034	--TWPE-----	1063
Db	1115	GTTTKPGIPVNGGVVPDEEAKQADKQGLIVPPTTNSINKDPVTNTQYSNTTGN---	1172
Qy	1064	VEYNPKSEDAGNEGETPHMLLRPHFVPEVTPD-----MDYLPVNPQOIGIINP---	1113
Db	1173	--INP--ETGKVIPLSGSLNPSYFNTPOQTDEITGPKVDVTGTPYDPSTGEIIDPAT	1228
Qy	1114	-----MLSDENYCNCKG-----	1139
Db	1229	KLPIPGSVAGDEILTEVLNITTDVETGLPDTLETGLPRDPVSGLPQGLPNTGLV	1281

RESULT 7  
US-07-638-431-2  
: Sequence 2, Application US/07638431  
: Patent No. 5198535  
: GENERAL INFORMATION:  
: APPLICANT: Hoffman, Stephen L.  
: APPLICANT: Charcoenvit, Yupin  
: APPLICANT: Hedstrom, Richard  
: APPLICANT: Khumsmith, Srisin  
: APPLICANT: Rogers IV, William O.  
: TITLE OF INVENTION: Protective malaria sporozoite surface protein  
: NUMBER OF SEQUENCES: 2  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: A. David Spevack  
: STREET: NMDC Building 1 T-12 National Naval  
: STREET: Medical Center  
: CITY: Bethesda  
: STATE: MD  
: COUNTRY: USA  
: ZIP: 20814-5044  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.24  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/07/638,431  
: FILING DATE: 19910110  
: CLASSIFICATION: 424  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Spevack, Avrom D.  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (301) 295-6759  
: TELEFAX: (301) 295-4033  
: INFORMATION FOR SEQ ID NO: 2:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 826 amino acids  
: TYPE: AMINO ACID  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: IS-07-638-431-2

[illegible]

[illegible]



```

: Sequence 5, Application US/08700651B
: Patent No. 6015882
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUT, JIRI
: TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
: TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
: FILE OF INVENTION: INFECTIONS
: FILE REFERENCE: 480.19-4(HV)
: CURRENT APPLICATION NUMBER: US/08700, 651B
: CURRENT FILING DATE: 1997-08-14
: EARLIER APPLICATION NUMBER: 08/415, 751
: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 1721
: TYPE: PRT
: ORGANISM: Cryptosporidium parvum
: US-08-700-651-5

```

[illegible]

```

608  --TTLNPIITTTTQKQRTTTT-----TT-----KVQKPPAIATTTTTLKPIVITTTTKATT 656
Qy 677 PKETAPTTKGTAPTTLKKEPAPTTPKKPAK-----ELAPTTTKEPT-----719
Db 657 TTTTVPT-----TTTTRKDEMTTTTTLPLDIDGIEITPPIEKMLDKYTRMYIYOVNSGL 711
Qy 720 --TTSOKPAPTTPKG--TAPTTPKKEPAPTTKEPAPTTPKGTAPTTKE-----PAPTTPKPK 773
Db 712 LLDNSDEPIPGSOAGIADTSLNLFVQVTHKSTGLPIDPVMVGLFDPKSGNLVHPYNTQM 771
Qy 774 APKELAPTTTKGPTSTTSOKPAPTTPKETAPTTKEPAPTTPKKAPPTPETPPPTSEV 833
Db 772 SGLSVSYLAAKNLTVDTDETGY--LPIDTLGYPLDPVSLIPN--PETGELEDPIDSDBI 827
Qy 834 STPT-----TTKEPTTIHK-----PDEST- 853
Db 828 MNGTAGIVSGISASESLLSOKSALIDPATNMVVGEFGLLNIPATGMVPIPGFLGSPSEQ 887
Qy 854 -----PELSAEPPTKALENSPKPGVPPTTKTAAATKPEMTTTAKDKTTTER--- 898
Db 888 FSPETEDGIIIPPEVAANAADKELSLIP--PSVP-----ESIP-----KQKIDISE 934
Qy 899 -----DLRPTTETTTAAKPMKETAATTEKTTESKITATT--TQVSTTTTQD 943
Db 935 LMWDIESRLTGQVSKRPIPGSIAGDLNPIIMKTPTQDTSVTGKPIDPTTGLPFNPPTGHL 994
Qy 944 TTPPKITLTKTTTLAPKVTTTKTITTTTEIMNKP--EETAKPKD-----RATNSKATT 994
Db 995 INPTNNNTMDSFAGAYKAVSNGIKTDNVYGLPVGEITGLPKDPGSDIPNSTTIGELVD 1054
Qy 995 PKQKPTKAPKKPTSTKKPKTMP-----RVRKPTTTPTRPKMWS 1033
Db 1055 PSTGKPINNSTAGIVSGKGLPIEDENGLNLPDPSNLPIDGNLNOLVNPETNSTVSGSTS 1114
Qy 1034 --TMPE-----LNPTSRIAEAMLQTTTRPNOTPN SKL 1063
Db 1115 GTTPKPGPIPVNGGVVPDDEAKDQADKKGGLIVPPTNSINKDPVTNTQVNTTGNL-- 1172
Qy 1064 VEVAPKSEDAGGAEGETPHMLLRPHVMEVTPD-----MDYLRVPHQGGIIINP-- 1113
Db 1173 --INP--ETGKVIPIGSLPGSLNVPSENTPQQTDEITGPKPVDVTGLPYDPSTGEIIDPAT 1228
Qy 1114 -----MLSDETNICNGK-----PVDGLTTLNRNGLV 1139
Db 1229 KLPIPGSVAGDEILTEVLNITTDVTGLPIDLETGLPRDPVSGLPQLPNGTLV 1281

RESULT 5
US-09-103-429A-3
; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: CDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th Tlaga
; CITY: IChaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998

```

us-09-556-246-1\_copy\_25\_1404.ra

Mon Apr 29 08:35:42 2002

552 ----- 551  
Db MIPNDTHVRFRKVKDVGNTISVRCKGAGKLEFFDRSLDTIPPVAGHNSCIIIVGS 629  
552 -----PAPTPKE-----PAPTPKEPAPTTTKK 575  
Db GGGKIHVSYPGSKDVLISALIOQCELFNEVYCDCTAKYGAHSGVTSADFVTTTAK 689  
576 PAPAPKEPAPTTKETAPTTPKLTTPTEKLAAPTEPEAPTTPEELAPTTPEEPTPT 635  
Db PTTTT--TGAPCQPTTTTGSPSAPTTTTTKATTT-----TILNPIITITQKPTT 741  
576 TPEEPAPTTKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 695  
Db T-----T-----KVPKPIATTTTLKPIVTTTTTKATTTTTPPT-----TPTTKR 787  
596 PAPTPPKKAPK-----ELAPTTKEPTS-----TTSKAPAPTPKG--TAP 735  
Db DEMTTTTPLDIDIGIETPIPIEKMLDKYTRMIYDYNGLLSDNDEPIPGSQAGIAD 847  
736 TTPKEPA-----PTTPKEPAPTPK----- 755  
Db TNLNLPVQTHKSTGLPIDPMVGLPDPKSGNLVHPYTNMTMGLSVSYLAAKNLVDIDE 907  
756 --GTAPTTLK--PPAPTPPKKAPKELAPTT-----KGTSTTSKDP 794  
Db TYGLPIDTLTGYPDLPSLIPFNPETGLFDPISDEIMNGTIAGIVSGISASELLSOKS 967  
795 APPTKE-----TAPTTKEPAPTTTPKAPTTPEP-----PPTSEVSTPTT 838  
Db APIDPATNMVVEFGLLNPATGVMLPGSLGSEQTPSPETEDGIIIP-----EVAANA 1024  
839 TKPPTTHKSPDESTE-----LSAEPTRKALENSKPEKGPV 875  
Db DFKLSIPSPVPSIPEKDKIDSISELMYDIESGRIGQVSKRPIPGSIAGDLN----P 1080  
876 TTKPATKPEMTTAKDKTTERDLRTTETTTAAKMTKETA-----TTTEKTTESKITA 931  
Db IMKTPTQT-----DSVTGKPI-----DPTTGLP--FNPTGHLINPNNNTMDSFAG 1126  
932 TTTQVSTTTQDTPPKITTLKTTLAPK--VTTTKKTTTTEIMMKPEETAKPKDRAT- 988  
Db AYKAVNSGIKTDVYGLPVEITGL--PKDPVSDIPFNSTTGLVD--ESTGKPINNYTA 1183  
989 -----NSKATTPKQKAPKAPKPTSTKKPKTPVRVRKPKTTPTRKMTS 1033  
Db GIVSGKRGLPIEDENGLNFDPTKULPIDGNLNVNPNSTVSGSTSGTKPKGIPVN 1243  
1034 --TMPE-----LNPTSIAEAMLOTTTRPNOTPNKLVNPNKSEADAG 1075  
Db GGVVPEEAKDQADKGDGLIVPPTNSINKDPVTNTQYSNTTGN-----INP--ETGKV 1297  
1076 AEGETHMLLRPHVMEVTPD-----MDYLPVYPNOGIILNP----- 1113  
Db IPGSLGSLNPSFNTPEQTDITKGVDAVDTGLPYDPSTGEIDPATKLPIDGVSAGDE 1357  
1114 MLDNETNLCKG-----PVDGLTTLRNGTLV 1139  
Db ILTEVLNITTTDEVTLGLPIDLETGLPDPVSLPQLPGLNGTLV 1398

RESULT 3  
520236-25  
; Patent No. 520236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 25-MAY-1990  
; APPLICATION NUMBER: 82,456

; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO: 25  
; LENGTH: 744  
520236-25

Query Match 6.8%; Score 506.5; DB 6; Length 744;  
Best Local Similarity 29.9%; Pred. No. 6e-24;  
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 226 PKITTTAKPINRPSLPNSDTSKETSLSLVNKETTVETKETTNTKQSTDGEKETSKE 285  
Db 24 PKMTYPTPKPSYPP-----TVKSKPTY-----PKKIT--- 53  
QY 286 TQSIKTSKADLAPTSKVLAKTPKAEITTKGPALTTPKEPTTPKE-----PASTPKE 341  
Db 54 -----YPTTYK--AKPS-----YPTTYKPKKTYPTPKKLTYPPTPKKLTYP 92  
QY 342 PTTTTKSAPT--TPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE--P 398  
Db 93 SYPTTKSKETTYKPKTYPTPKKPSYPTPKKTYPTPKKTYPTPKKTYPTPKKTYPTPK 151  
QY 399 TTTKSAPTTPKEPAPTTTPKAPPT--TPKEPAPTT-----PKEPTTPKEPAPTTKEP--- 450  
Db 152 PTYKPKSYPP--PSYTKTKTYPTPKKLTYPPTPKKPSYPTPKKPSYPTPKKLTYP 209  
QY 451 -APTTKEPA--PTAPKKA--PTTPKEPA--PTTPKEPA--PTTPKEPS--PTTPKEPA 499  
Db 210 YPTTYKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPK 269  
QY 500 -PTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 552  
Db 270 YPSYKAKPSYPT--YPTSYKAKPSYPT--PTYKAKPSYPTPKKTYPTPKKTYPTPK 322  
QY 553 A--PTTPKEPA--PTTPKEPA--PTTPKKA--PTAPKEPAPTTTPKEPAPTTTPKLTPTT 604  
Db 323 SYPTTKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPK 382  
QY 605 DEKLAPTPPE--KPAPTTPPELAPTTPEEPT--PTTPKEPA--PTTPKEPA--AAPT---P 653  
Db 383 TYKAKPSYPTPKKPSYPT--PTYKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPT 438  
QY 654 KEAPAPTTTPKEPAPTT-----PKEPAPTTTPKEPAPTT--PKGTAPTTPKEPAPTTTPK 707  
Db 439 SYPTTKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPK 498  
QY 708 ELAPTTTKEPTSTTSKAPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPA- 766  
Db 499 SYPTTKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPK 546  
QY 767 -PTTPKKAAPKELAPTTTKGPT--STTSKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 821  
Db 547 PPTTKKAPTYK--AKPTTKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPT 601  
QY 822 ---TPETTPPTTSEVSTPTTKE-----PTTIHKSPTD-----ESTPELSAEP--PKALENS 868  
Db 602 YKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPT 661  
QY 869 PKEPGVPTTKTAA-----TKPEMTTAKDKTTERDLRTTPTTAAAP-----KMT 914  
Db 662 PSYPP--PTYKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPT 719  
QY 915 KETATTTTEKTESKITATTT 934  
Db 720 YKAKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPTPKKPSYPT 739

RESULT 4  
US-08-700-651-5

Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

QY 300 TSKYLAKPTPKAETTTKGPALTTKPEPTTTKPEASTTKPEPTTTTKSAP-----TTP 354  
DQ 9 TALGLVAARPEVSDAEKNFALHEPHDPDX--PAEQXXLLPXDEYCTKPYCYEYGLKFIAP 66  
QY 355 KEAPATT-----TKSAPTTKPEAPTTTTPKPEATTTPKPEATTTPKPEATTTK 402  
DQ 67 RDCAPGTGKFSAQTCVHAALAGCTLPGPPAETT---QAPATTO--APTITQ--APTITQ 119  
QY 403 SAPTTKPEAPTTPKKPARTTKPEAPTTKPEPTTTKPEAPTTKPEATTTPKPEAPT 462  
DQ 120 QAPTTTTQATTT-----QAPTTQAPTTT---QATTTQAPTTT-----156  
QY 463 PKPARTTKPEAPTTKPEAPTTTKPESTTPKPEAPTTTKSAPTTTKPEATTTKSAP 522  
DQ 157 ---QAPTTTQ--APTITQ--APTITQ-----APTITQ--APTITQ--APTITQAA 198  
QY 523 TTPKPESTTPKPEAPTTTP--KEAPTTPKPARTTKPEAPTTTPK-----566  
DQ 199 TTPAATTTPAAT--TPAATTTPAATTTPAATTGVPAPTS---APVMPPICELLPNCPADFD 253  
QY 567 -----EPAATTTK 574  
DQ 254 IHLIPHDKYCNLFQCSNGYTFEQRCEGLYFNPYQRCDSANVECDGEISPAVPVTE 313  
QY 575 K-----PA-----577  
DQ 314 GNEDEDIDGILLONGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCGAGTHFSFELQ 373  
QY 578 -----PTAP--KEAPT-----TPKET 592  
DQ 374 QCDHIELVGCTLPGSESEVDVEDACTGWYCTEPLEWEPLPNCGPADFSDHLLPHES 433  
QY 593 -----APTPKKLLPTT-----PEKLAPTTPEKAP 618  
DQ 434 DGGVYQCVHGQT IARPCPNLHFSPATQCESPVTAGQVFCEDSDNOCTSTAAPTAP 493  
QY 619 TTPBELAPTTPEEPTTPPEEAPTTPKAAANTPKPEAPTTTPKPEAPTTTPKPEAPTTPK 678  
DQ 494 TAAPTAAPTAAPTAAPSTVVPPA--TPATAAPVPTTAIPT-----PAPTAAPTAAPTAA 548  
QY 679 ETAPTT---PKGTAPTLKPEAPTTPKKPAKELAPTTTKPTSTSDKPAATTPEKGTAP 735  
DQ 549 PESPTTVTPPTAAPT-----APTT-----AVPEIPTVSAPTAAPT--AAPTAPTAAP 598  
QY 736 TTPKPEAPTTTPKPEAPTTKGTAPTTKPEAPTTKPKPAKELAPTTTKGPTST--TSDKP 794  
DQ 599 TTAVPELPTT-----VTSPPPTAAPT--AAPANT-----TAVTPPTAAPTAA 641  
QY 795 APTTPKETAPTTKPEAPTTTPKPAATTTPETPPPTTSEVSTPTTTKPTTHKSPDESTP 854  
DQ 642 AP-----NTTVPPTAAPT--AAPANTTVTPPTAAPTAP-----PTVAH-----AP 685  
QY 855 ELSAETPKALENSKPEPGVPTTKTPAATKPE 886  
DQ 686 NTTAAP-----VTTTSAPATT--PE 703

RESULT 2  
US-08-928-361B-5  
; Sequence 5, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS.  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENT  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6

Mon Apr 29 08:35:42 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:16:35 ; Search time 49.78 Seconds  
(without alignments)  
623.836 Million cell updates/sec

Title: us-09-556-246-1\_COPY\_25\_1404

Perfect score: 7410  
Sequence: 1 QDLSSCAGRCGEGYSRATC.....ARAITRSQTLKVKVYNCP 1380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	522	7.0	805	4	US-09-103-429A-4
2	513	6.9	1837	3	US-08-928-361B-5
3	506.5	6.8	744	6	5202236-25
4	499.5	6.7	1721	3	US-08-700-651-5
5	498.5	6.7	1721	3	US-09-103-429A-3
6	498.5	6.7	1721	3	US-08-928-361B-6
7	488.5	6.6	826	1	US-07-638-431-2
8	488.5	6.6	826	1	PCT-US92-00018-2
9	476.5	6.4	652	6	5202236-13
10	452	6.1	960	4	US-09-219-849-5
11	425.5	5.7	1185	4	US-09-041-886-23
12	424.5	5.7	1867	2	US-08-479-537A-5
13	424.5	5.7	1867	4	US-09-083-116-5
14	424.5	5.7	2035	2	US-08-479-537A-2
15	424.5	5.7	2035	4	US-09-083-116-2
16	419.5	5.7	2476	2	US-08-276-967-2
17	417	5.6	829	1	US-08-642-255-132
18	417	5.6	829	1	US-08-397-633A-53
19	417	5.6	837	1	US-08-175-155-68
20	417	5.6	837	1	US-08-477-509B-103
21	417	5.6	837	1	US-08-642-255-101
22	417	5.6	837	2	US-08-707-237A-75
23	417	5.6	837	3	US-08-482-085B-103
24	417	5.6	897	1	US-08-397-633A-50
25	413.5	5.6	907	3	PCT-US95-04611A-19
26	413.5	5.6	907	5	PCT-US95-04611A-19
27	404.5	5.5	408	1	US-07-609-716-65

28	404.5	5.5	408	4	US-08-475-411A-65	Sequence 65, Appl
29	404.5	5.5	408	4	US-08-478-029A-65	Sequence 65, Appl
30	398.5	5.4	682	1	US-08-642-255-126	Sequence 126, App
31	398.5	5.4	682	1	US-08-397-633A-36	Sequence 36, Appl
32	384	5.2	1537	1	US-08-325-267A-2	Sequence 2, Appl
33	383.5	5.2	1231	3	US-08-904-263A-4	Sequence 4, Appl
34	381	5.1	1848	4	US-08-296-791-6	Sequence 6, Appl
35	381	5.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
36	376	5.1	960	4	US-09-219-849-6	Sequence 8, Appl
37	373	5.0	1786	4	US-08-973-462-8	Sequence 6, Appl
38	371	5.0	761	2	US-08-707-237A-84	Sequence 84, Appl
39	371	5.0	762	1	US-08-642-255-114	Sequence 114, App
40	371	5.0	762	1	US-08-397-633A-26	Sequence 26, Appl
41	370.5	5.0	1064	1	US-08-642-255-62	Sequence 62, Appl
42	370.5	5.0	1187	1	US-08-320-559-28	Sequence 28, Appl
43	370.5	5.0	1187	3	US-08-545-860D-28	Sequence 28, Appl
44	370.5	5.0	1187	5	PCT-US94-04496-28	Sequence 28, Appl
45	370.5	5.0	1210	1	US-08-320-559-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1  
US-09-103-429A-4  
; Patent No. 6187558  
; Patent No. 6187558  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 No. 6187558th Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/09/103.429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 805 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic membrane  
; US-09-103-429A-4

Query Match 7.0%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 7.2e-25;

\_\_\_\_\_

Mon Apr 29 08:35:43 2002

Db 958 KEEKAKK-----PEEKPKTEAKAKEDDKTISKPSKPKAEKAKSSSTDQKDSKPPKAT 1012

DR PIR; S00979; QFHUH.  
DR MIM; 162230; -  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.  
FT DOMAIN 1 100 HEAD.  
FT DOMAIN 101 413 ROD.  
FT DOMAIN 414 1020 TAIL.  
FT DOMAIN 101 132 COIL 1A.  
FT DOMAIN 133 145 LINKER 1.  
FT DOMAIN 146 244 COIL 1B.  
FT DOMAIN 245 266 LINKER 12.  
FT DOMAIN 267 288 COIL 2A.  
FT DOMAIN 289 292 LINKER 2.  
FT DOMAIN 293 413 COIL 2B.  
SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCFLD4 CRC64;

Query Match 6.8%; Score 502; DB 1; Length 1020;  
Best Local Similarity 26.7%; Pred. No. 3.4e-13;  
Matches 209; Conservative 81; Mismatches 356; Indels 136; Gaps 35;

QY 119 KTKKVIIESE--ETFEHSVS-ENOESSSSSSSSSSSTIWKIKSSKNSAANRELQKLLK 175  
DB 336 KSTKDSLERQSELEDRHQADIASYQEAIOQLDAELRNTKWM-----AAQLREYQDILLN 390  
QY 176 VK-----DNKKRTKKKPTKPPVVDEAGSLDNGDFKVTTPDTSTTOHKN 221  
DB 391 VKMALDIEIAAYKLLGECEGRIGGPID-----FSLPEGLPKI--PSVST--HIK 437  
QY 222 VSTSPKITTAKPINRPSLPPNSDTSKETSLLVKNKETTETKTTTNKQISTDGKE-KT 280  
DB 438 VKSEEEKIKVVE-----KSEKETIVVE-----QTEETQVTEEVTEEEKEAKE 480  
QY 281 TSAKETQSIEKTSKDLAPTISKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPASTTPK 340  
DB 481 EEGKEEGEGEEAEAG-----GEEETKSP-----PAEEAASPEKEAKSPVKE 522  
QY 341 EPTPTTIKSAPTTTPKEPATTTKKSAPTTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 400  
DB 523 EAKSPAEEKSPKEEAKSPAIVKS-PEKAKSPA-----KEEAKSPPEAKSPEKEEAKSPA 577  
QY 401 TKSAPTTTPKEPATTTPKAPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPAP 460  
DB 578 VKS-PEKAKSPAEEAKSPA-----EAKSPEKAKSPVKEEAKSPA-EAKSPVKEEAKSPA 631  
QY 461 TAPKAPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKSAPTTTPKEPATTTPKS 520  
DB 632 VKSPEKAKSPKEAKSPEKAKSPEKAKSPEKAKSPEKAKSPEKAKSPEKAKSPEKAKS 690  
QY 521 APTTPKEPSPTTTPKEPATTTPKEPATTTPKAPATTTPKEPATTTPKEPATTTPKAPATA 580  
DB 691 -PEKAKSP----VKEEAKSPEKAKSPVKEEAKSPEKAKSPV-----KEEAKTPKAKSPV- 740  
QY 581 PKEPATTTPKETATTTTPKLTTPPEKLTTPPE--KPAPTTPPEEL-APTTPPEEPTTTP 637  
DB 741 -KEEAKSPEKAKSPEKAKTLDVKSPEAKTFAKEARSFADKFEKAKSPVKEEVKSPKEA 799  
QY 638 EEPATTTPKAAAPNTTPKEPATTTP-----KEPATTTPKEPATTTPKETATTTTPKGTATTL 693  
DB 800 KSPKADAKAPEKEIPKKEEVKSPVKEEKPQEVKVPKPEKAPATPK-----852  
QY 694 KEPATTTPKAPKELAPATTTPKETSITSDKAPATTTPKGTATTPKEPATTTPKEPATTTP 753  
DB 853 TEEKKDSKKEAPKKEAP-----KPKVEEKKEPAVEKESKVEAKKEAEADKKVP--T 905  
QY 754 PKGTAP--TTLKEPATTTPKAPKELAPATTTPKGTSTSDKPA-PTTPKETATTPKEP 810  
DB 906 PEKAPAKVEVKEDA-----KPKETEVAKEPDDAKAKPEKSPKAEKKEAAP-EKKDT 957  
QY 811 APTTPKAPATTTPPTTPPTTSEV--STPTTTPKEPTTIHKSPE--STPELSAPPTPKAL 865

QY 866 EN 867  
DB 1013 ED 1014  
Search completed: April 26, 2002, 16:19:21  
Job time: 346 sec

[2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE-95347616; PubMed-7622068;  
 RA Mita K., Tsuji H., Morimyo M., Takahashi E., Neno M.,  
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;  
 RT "The human gene encoding the largest subunit of RNA polymerase II.";  
 RL Gene 159:285-286(1993).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X63564; CA44125.1; .  
 CC EMBL; X74874; CA52862.1; .  
 CC EMBL; X74873; CA52862.1; JOINED.  
 CC EMBL; X74872; CA52862.1; JOINED.  
 CC EMBL; X74871; CA52862.1; JOINED.  
 CC EMBL; X74870; CA52862.1; JOINED.  
 CC EMBL; X74870; CA52862.1; JOINED.  
 CC MIR; S21054; S21054.  
 CC MIM; 180660; .  
 CC InterPro: IPR000684; RNA\_polII\_repeat.  
 CC InterPro: IPR000722; RNA\_polA.  
 CC InterPro: IPR002879; RNA\_polA2.  
 CC Pfam: PF00623; RNA\_polA; 1.  
 CC Pfam: PF01854; RNA\_polA2; 1.  
 CC PROSITE: PS00115; RNA-directed II REPEAT; 43.  
 CC Transferrase; DNA-directed RNA polymerase; Transcription; zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.  
 CC ZN\_FING 71 87 C2H2-TYPE (POTENTIAL).  
 CC DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC CONFLICT 1067 1067 W -> L (IN REF. 2).  
 CC CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 CC SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 6.8%; Score 503.5; DB 1; Length 1970;  
 Best Local Similarity 33.3%; Pred. No. 5.1e-13;  
 Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;  
 QY 326 PTTTTPKEASTTPE-PTPTTKSAPTTPKEPAPTTKAPTTPKE-----PAPTTTKE 379  
 DB 1507 PSPMGGISPAWTPWNGATPAYGAWSPSVGSGWTPGAAAGFSASADSGFSGSPAWS 1566  
 QY 380 PAPTTPKEPAPTTKAPTTKSAPTTPKEPAPTTKAPTTKAPTTKPE--APTTPKEPTP 437  
 DB 1567 PTPGSGRGSPSPYIPSPGGAMSPSYSPSPA-YEPRSPGTYTQSPSYSPSPSY-SP 1624  
 QY 438 TTPKEPAPTTKPEPAPTTKPEAPTKAPTKAPTTKPEPAPTTKPEPAPTTKPEP 497  
 DB 1625 TSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1671  
 QY 498 PAPTTPKSAPTTPKEPAPTTKSAPTTPKEPAPTTKPEPAPTTKPEPAPTTKPEP 557  
 DB 1672 -SPTSPSYSPSPSY-SPTSPSYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1725

QY 558 KEPAPTTKPEPAPTTTTPKAPTAPEKAPAPTTKAPTTKAPTTKAPTTKAPTTKAPTA 617  
 DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-T 1777  
 QY 618 PTTPEELAPTTPEPTTTPPEAPTTKAAANTPKAPAPTTKPEPAPTTKPEPAPTTKPEP 675  
 DB 1778 PTPS-NYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1828  
 QY 676 TPKEPAPTTKPGTAPTTKPEPAPTTK-KAPKELAPTTTKEPTSTSTSDKAPAPTTKGT 734  
 DB 1829 SP-SYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY 1875  
 QY 735 PTPKPEPAPTTKPEPAPTTKPGTAPTTKPEPAPTTKPKAPKELAPTTKGT-SPTSDK 793  
 DB 1876 PTPSTY-SPTTPKY-SPTSP-SYSPSPSY-SPTSP-SYSPSPSY-SPTSP-SYSPSPSY 1922  
 QY 794 PAPTTPKETAAPTTPKEP--APTTPKAPAPTTPEPTTSEVSTPTTPPTTKEPTTHKSPDE 851  
 DB 1923 YSTSP-TYSPSTPKSTYSPTSPGY-SPTSP-----TYSLTSPAI-----SPDD 1965  
 QY 852 STPE 855  
 DB 1966 SDEE 1969  
 RESULT 15  
 NFH\_HUMAN  
 ID NFH\_HUMAN STANDARD; PRT; 1020 AA.  
 AC P12036;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).  
 GN NEFH OR NEFH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-88328981; PubMed-3138108;  
 RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;  
 RT "The structure and organization of the human heavy neurofilament  
 RT subunit (NF-H) and the gene encoding it.";  
 RL EMBO J. 7:1947-1955(1988).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X15306; CAA33366.1; .  
 CC EMBL; X15307; CAA33366.1; JOINED.  
 CC EMBL; X15308; CAA33366.1; JOINED.  
 CC EMBL; X15309; CAA33366.1; JOINED.

DR	PROSITE: PS00025; P-TREFOIL; 6.	Query Match	6.9%;	Score 509.5;	DB 1;	Length 662;
KW	Repeat; Amphibian skin; Glycoprotein; Alternative splicing.	Best Local Similarity	28.1%;	Pred. No. 1.2e-13;		
FT	NON_TER	Matches 223;	Conservative	48;	Mismatches 276;	Indels 247;
FT	DOMAIN	81	144			Gaps
FT	REPEAT	81	88			
FT	REPEAT	89	96			
FT	REPEAT	97	104			
FT	REPEAT	105	112			
FT	REPEAT	113	120			
FT	REPEAT	121	128			
FT	REPEAT	129	136			
FT	REPEAT	137	144			
FT	DOMAIN	161	202			
FT	DOMAIN	218	301			
FT	REPEAT	218	224			
FT	REPEAT	225	239			
FT	REPEAT	240	249			
FT	REPEAT	250	259			
FT	REPEAT	260	275			
FT	REPEAT	276	287			
FT	REPEAT	288	294			
FT	REPEAT	295	301			
FT	DOMAIN	306	347			
FT	DOMAIN	353	394			
FT	DOMAIN	402	522			
FT	REPEAT	402	411			
FT	REPEAT	412	419			
FT	REPEAT	420	431			
FT	REPEAT	432	443			
FT	REPEAT	444	453			
FT	REPEAT	454	460			
FT	REPEAT	461	472			
FT	REPEAT	473	479			
FT	REPEAT	480	491			
FT	REPEAT	492	498			
FT	REPEAT	499	515			
FT	REPEAT	516	522			
FT	DOMAIN	525	566			
FT	DOMAIN	572	613			
FT	DOMAIN	620	661			
FT	DOMAIN	162	188			
FT	DISULFID	172	187			
FT	DISULFID	182	199			
FT	DISULFID	307	333			
FT	DISULFID	317	332			
FT	DISULFID	327	344			
FT	DISULFID	354	380			
FT	DISULFID	364	379			
FT	DISULFID	374	391			
FT	DISULFID	526	552			
FT	DISULFID	536	551			
FT	DISULFID	546	563			
FT	DISULFID	573	599			
FT	DISULFID	583	598			
FT	DISULFID	593	610			
FT	DISULFID	621	647			
FT	DISULFID	631	646			
FT	DISULFID	641	658			
FT	VARIANT	276	276			
FT	VARIANT	354	354			
FT	VARIANT	415	415			
SQ	SEQUENCE	662 AA;	67774 MW;	F085277F1ED2FD40	CRC64;	

Query Match

Best Local Similarity

Matches 223; Conservative

6.9%;

Score 509.5;

DB 1;

Length 662;

28.1%;

Pred. No. 1.2e-13;

48;

Mismatches 276;

Indels 247;

Gaps

Qy	267	TTNKQFSTDGKBEKTSKAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTPKEP	326
Db	3	TTAAVAATGKDTTAAAGCSAAAEKTA-----AGEVSAPPT--AAVATGEDATT-----	51

QY	327	TPTTPKEPASTTTPKEPTPTTIKSGAPTTTPKEPATTTTTKSAPTTPKPEAPTTTKPEAPTTTPKEAPTTTPK 386
Dd	52	: : : : : AAATAAEITAAAGEAPTTTTATAPTAAGAKARTTAAATAPTAAAGAPTAT 103
QY	387	EPARTTTTKEPATTTTKS-APTTPKEAPATTTPKKPAPTTPKPEAPTTTPKEAPATTTPKEAPTTTPKEA- 444
Dd	104	GKAPATAAAPVPTTTAAASKAPTTHSTAAAAHTSTAAAAAPTTHSTAASAKSKERSTSSSEEEHCH 163
QY	445	-PTTKKEAPN---TPKEAPTAAPKKPAPTTPK-----EPARTTPKEAPTTTPKE 489
Dd	164	VKPSAREMGSGKITKKQ----CKKNCCFDPKGGHGHCIFHRKPKGHSHEHHTTTTK- 218
QY	490	PSPTTPKEPATTTTKSAPTTPKPEAPTTTKSATPTTPKEPSPTTPKEAPTTTPKEAPTTTPKEAPTTTP 549
Dd	219	-----APTIIQIAHTTT---TPPTTT-----TTTTKATPTT----- 245
QY	550	KKPAPTTPKEPATTTTPKPEAPTTTKKAPTAPKPEAPTTTPKETARTTPKKLTPTTPKEKLA 609
Dd	246	-----TTTTKATPTT-----TTTTKATTTT---TPTT----- 270
QY	610	P TTPEKPATTTPDEELAPTTPDEPTTPTEBPAPTTPKAAPNTTPKEAPTTTPKEAPTTTP 669
Dd	271	-----TTTTTKATTTTPTTTTTP----- 289
QY	670	KEPATTPKETAPTTPKGATPTTLKEAPATTTPKKPAPKELAPTTTKEPTSTTSDKPAPT 729
Dd	290	-----TTTTTKATTTTTSGECKMEPSK-----RDCGYSGITESQCR 328
QY	730	PKG-----TAPPT-----PKEPATTTPKEAPATTTPKGATPTTLKEAPATTTPKKPAPKELA 779
Dd	329	TKGCCDFDSSIPOTKWCFYTLSOAVDKVEPSQRVDCCFGRGIT-----ADQCQNCCFDSS 384
QY	780	PTTTKTGTSTTSDDKAPATTTPKETAPTTPKEPATTTTPKKPAPTTPPETPPPTTSEVSTPTTT 839
Dd	385	ISGTWKCFYSTSOVA--TKTTTTPTTTTTTTTTTKATTTTP-----TTTTTTPTTT 436
QY	840	KEPTTIHKSPDESPELSAETPKALENSPKPGVPVTIKTPAAATKPEMTTAXDKDTERD 899
Dd	437	TTTTT-----TTKA-----TTTTTTPTTTTT---KAT--- 463
QY	900	LRTTPTTTAAPKMTKETATTTTEKTESKITATTTQTVSTTTODTTPFKITTLTKTTLAP 959
Dd	464	-TTTTPTTTTTP-----TTTTTKAT-----TTTTTTTTTTTTTTTKATTTPTTTTT 510
QY	960	KVTTTKKTTTITTIMKNKPETAAPKDRATNSKATTPKQKTPAKPKPTSTKKPK-----T 1015
Dd	511	TTTTTKATTTTTTSGECKME-----PSKRADCYGPIGITESQCRSGCCPDFSSIPQTKWCFYS 566
QY	1016	MPRVRKPKTTPTPR 1029
Dd	567	LPOVADCKVAPSSR 580
RESULT	14	
ID	RBP1_HUMAN	STANDARD; PRT; 1970 AA.
AC	P24928;	
DT	01-MAR-1992 (Rel. 21, Created)	
DT	01-MAR-1992 (Rel. 21, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=92178992; PubMed=1542581;	
RA	Winterlitz M., Acker J., Vicaire S., Vigneron M., Keding C.;	
RT	"Complete sequence of the human RNA polymerase II largest subunit."	
RL	Nucleic Acids Res. 20:910-910(1992).	

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=1542581;  
 RA Wintzerith M., Acker J., Vicaire S., Vigneron M., Keding C.;  
 RL "Complete sequence of the human RNA polymerase II largest subunit.";  
 RT Nucleic Acids Res. 20:910-910(1992).



```

Db      479 SPAGAKPPPAKSPAEAKSPAEEKSPEVVKSPKSAKSPVKEGAKSLAE   538
QY      521 --APTTPKEPSPTTTKEPA---PPTPKEPARTTTPKKPARPTTKPEAPRTTPKKEPARTTK   574
Db      539 AKSPEKAASPVKEIKRPAEAVKSPKSAKSPMRKEAKSPEKAKTLDVKSPEAKPPAKEEK   598
QY      575 KPAPTAPKEPARPTTPKETARTTPKKLTPTTPTEKLAPRTTPKEKPARTTPEELAPRTTPEERTP   634
Db      599 RADIIRSPQVQSPAKBEAKSPEKEETRT--EKVAPKKEEVKSPV--EEV---KAKEPPK   651
QY      635 TTPPEEAPRTTPKAAANPTPKEPARTTPKKEPARPTTPKPEAPRTTPKETARTTPKGTARTTTLK   694
Db      652 KVEEBKTATPKTEVKESKKDEAPKAQKP-KAEEKEPLTEKPKDSGGEAKKEEA---K   706
QY      695 EPARTTPKPAKELAPRTTTKETSSTTSDDKPARPTTPKGTAFTTPKPEAPRTTPKPEAPRTTP   754
Db      707 EKKAAPAEETPAKLGKVKEAKPKKAEDAQA-----KEPSKSEKE---NP   750
QY      755 KCTAPRTTLKEPARPTTPKPAKELAPRTTTKGTSTTSDDKPARPTTPKETARTTPKPEAPRTT   814
Db      751 K-----KEEVPAPAEKKDTKE-----EKTVESKKREEKPKMEAKAKEED----   789
QY      815 PKKPARPTTETPPPTTSETSVSTPTTTTKEPTTIHKSPDE   851
Db      790 --KGLPQLPEPSKPKTEKAEEKSSSTDQKDOSPSEKAPED   824

RESULT 13
MUC1_XENLA
ID    MUC1_XENLA          STANDARD;             PRT;       662 AA.
AC    Q05049;
DC    01-OCT-1994 (Rel. 30, Created)
DT    01-OCT-1994 (Rel. 30, Last sequence update)
DI    01-OCT-1994 (Rel. 30, Last annotation update)
DE    INTEGUMENTARY MUCIN C.1 (FIM-C.1) (FRAGMENT).
OS    Xenopus laevis (African clawed frog).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC    Xenopodinae; Xenopus.
OX    NCBI_TaxID=8355;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE-Skin;
RX    MEDLINE=93077556; PubMed=1447205;
RA    Hauser F., Hoffmann W.;
RT    "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT    C.1 (FIM-C.1) from xenopus laevis. Polydispersity and genetic
RL    polymorphism.";
RJ    J. Biol. Chem. 267:24620-24624(1992).
CC    -I- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC    INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC    -I- SUBCELLULAR LOCATION: SECRETED.
CC    -I- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN
CC    MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC    -I- TISSUE SPECIFICITY: SKIN.
CC    -I- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC    -I- SIMILARITY: CONTAINS 6 P-TYPE (TFREFOIL) DOMAINS.
CC    -----
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----
DR      EMBL; L02115; AAA74725.1; -.
DR      PIR; A45155; A45155.
DR      HSRP; P04002; IWFA.
DR      InterPro; IPR000519; P_trefoil.
DR      Pfam; PF00088; trefoil; 6.
DR      SMART; SM00018; P; 6.

```

DR	InterPro: IPR002965; P_rich_extensn.	DR	IPR002965; P_rich_extensn.
DR	PRINTS; PR01216; ADHESIVEL.	DR	PRINTS; PR01216; ADHESIVEL.
DR	PRINTS; PR01217; PRICHEXTENS.	DR	PRINTS; PR01217; PRICHEXTENS.
KW	SIGNAL: Repeat; Hydroxylation.	KW	SIGNAL: Repeat; Hydroxylation.
FT	SIGNAL 1 20 POTENTIAL.	FT	SIGNAL 1 20 POTENTIAL.
FT	CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.	FT	CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.
FT	DOMAIN 21 41 NONREPEITIVE LINKER.	FT	DOMAIN 21 41 NONREPEITIVE LINKER.
FT	DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-P-[PST]-[ST].	FT	DOMAIN 124 872 TANDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-P-[PST]-[ST].
FT	DOMAIN 184 192 NONAPEPTIDE 1.	FT	DOMAIN 184 192 NONAPEPTIDE 1.
FT	DOMAIN 213 221 NONAPEPTIDE 2.	FT	DOMAIN 213 221 NONAPEPTIDE 2.
SQ	SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;	SQ	SEQUENCE 872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;
Query Match 7.0%; Score 518.5; DB.1; Length 872;			
Best Local Similarity 29.2%; Pred. No. 6.9e-14;			
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;			
Qy	194 VVDEAGSLDNGDRKVTPTDST---TQH-----NKVSTSPKIITAKPINRPSLPNS 244	Qy	194 VVDEAGSLDNGDRKVTPTDST---TQH-----NKVSTSPKIITAKPINRPSLPNS 244
Db	29 VVGSAYGASAGAYK-TLPGSHPYGSKHVVYKPNKIPT-PYI--SKKSYAPYKPKGY 84	Db	29 VVGSAYGASAGAYK-TLPGSHPYGSKHVVYKPNKIPT-PYI--SKKSYAPYKPKGY 84
Qy	245 DTSKETSLTVNKETVETKETTTTWNKQSTDQCEKTTTSAKETQSTEKTSAKDLATSKVL 304	Qy	245 DTSKETSLTVNKETVETKETTTTWNKQSTDQCEKTTTSAKETQSTEKTSAKDLATSKVL 304
Db	85 YPTKRYQPTGYGSKTNPYPIYKIAKLSYKAIKTTYPAYKAKTSYPPSYK-----HKIT 139	Db	85 YPTKRYQPTGYGSKTNPYPIYKIAKLSYKAIKTTYPAYKAKTSYPPSYK-----HKIT 139
Qy	305 AKPTPKAETHTKGPALITPKPT-PTTKPEPASTPTKPEPTTIKSAPTTKPEPAPTTPK 363	Qy	305 AKPTPKAETHTKGPALITPKPT-PTTKPEPASTPTKPEPTTIKSAPTTKPEPAPTTPK 363
Db	140 YPTTYK-----PKITYP--PTYKQKSPSYKPKTYPPTYK-----PKITYPPTYK 185	Db	140 YPTTYK-----PKITYP--PTYKQKSPSYKPKTYPPTYK-----PKITYPPTYK 185
Qy	364 SAPT--TPKEPAPT--TTKPEAPTTPKEPAPTTPKPAATTTKSAPTTTPKEPAPT--PKK 418	Qy	364 SAPT--TPKEPAPT--TTKPEAPTTPKEPAPTTPKPAATTTKSAPTTTPKEPAPT--PKK 418
Db	186 RKPSTYPYKPKATYPPTYKPKITYP---PTYKRPKPSYT-----PYKPKTTPPTPKPKI 236	Db	186 RKPSTYPYKPKATYPPTYKPKITYP---PTYKRPKPSYT-----PYKPKTTPPTPKPKI 236
Qy	419 PAPTTPKEPAPT---PKPTPTT--PKPEAPTTPKEPAPTTPKEPAPT--APKPKAPTTP 471	Qy	419 PAPTTPKEPAPT---PKPTPTT--PKPEAPTTPKEPAPTTPKEPAPT--APKPKAPTTP 471
Db	237 SYPSIYKPKASVSSYKSKKTPPTYKPKISYPTTPYKPKPSYP---PTYKPKVTTPPY 292	Db	237 SYPSIYKPKASVSSYKSKKTPPTYKPKISYPTTPYKPKPSYP---PTYKPKVTTPPY 292
Qy	472 KPEAPTTPKEPAPTTPKESPTTPKEPAPTTPKSAPT-----TTKPEAPTTPKSAPTTPK 526	Qy	472 KPEAPTTPKEPAPTTPKESPTTPKEPAPTTPKSAPT-----TTKPEAPTTPKSAPTTPK 526
Db	293 K-PKPSYPTPKITYPPT-YKPKPSYPTPKQKPSYPPYLYKSKSSYPTSYKSKKATYP 350	Db	293 K-PKPSYPTPKITYPPT-YKPKPSYPTPKQKPSYPPYLYKSKSSYPTSYKSKKATYP 350
Qy	527 EPSPTTTKEPA-PTTKPEPAPTTPKKP-APT-TPK-----EPAPTTPKEPAPTTPKKP 576	Qy	527 EPSPTTTKEPA-PTTKPEPAPTTPKKP-APT-TPK-----EPAPTTPKEPAPTTPKKP 576
Db	351 TVPKKITYPPTYKPKPSYPPSYKPKKTSYTPYKPKITYPPTYKPKPSYPPSYKPKKATYP 410	Db	351 TVPKKITYPPTYKPKPSYPPSYKPKKTSYTPYKPKITYPPTYKPKPSYPPSYKPKKATYP 410
Qy	577 APTAPKEPAPTTPKEAP-----TTPKKLTPTTPKEAPTTTPKPAATTPPEELAPTTEE 631	Qy	577 APTAPKEPAPTTPKEAP-----TTPKKLTPTTPKEAPTTTPKPAATTPPEELAPTTEE 631
Db	411 T-YKPKISYPTPKKASVSSYKSKKTPPTYKPKISYPTTPYKPKPSYPTPKPKITYP 469	Db	411 T-YKPKISYPTPKKASVSSYKSKKTPPTYKPKISYPTTPYKPKPSYPTPKPKITYP 469
Qy	632 PT-PTTPEEPAPTTPKAAAPNTPK-EPAPTTPKEPAPTTP-----KEPAPTTP--PKETAP 682	Qy	632 PT-PTTPEEPAPTTPKAAAPNTPK-EPAPTTPKEPAPTTP-----KEPAPTTP--PKETAP 682
Db	470 PTYKPKPSYPTPKPKITYPPTYKPKPSYPTPKQKPSYPPYLYKSKSSYPTSYKSKKATYP 529	Db	470 PTYKPKPSYPTPKPKITYPPTYKPKPSYPTPKQKPSYPPYLYKSKSSYPTSYKSKKATYP 529
Qy	683 TT--PKGTAPTTLKEPAPTTPKPAKELAPTTPKPTSTTSOKPAPT--TPKGTAPTTPK 739	Qy	683 TT--PKGTAPTTLKEPAPTTPKPAKELAPTTPKPTSTTSOKPAPT--TPKGTAPTTPK 739
Db	530 PTYKPKITYPPTYK-----PKPSYPPSYKPKTYPPTYKPKIRPYPTYKPKASYPPTYK 583	Db	530 PTYKPKITYPPTYK-----PKPSYPPSYKPKTYPPTYKPKIRPYPTYKPKASYPPTYK 583
Qy	740 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP-----KKPAKELAPTTPKPTSTTSOKPA 795	Qy	740 EPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP-----KKPAKELAPTTPKPTSTTSOKPA 795
Db	584 -----PKITYPPTYK-PKPSYPTPKQKPSYPPYLYKSKSSYPTAYKSKKTY 628	Db	584 -----PKITYPPTYK-PKPSYPTPKQKPSYPPYLYKSKSSYPTAYKSKKTY 628
Qy	796 PTT--PKETAPTTPKEPAPTTPK--KPAPTTPET-----PPPTTSEVSPNTTKPEPT 844	Qy	796 PTT--PKETAPTTPKEPAPTTPK--KPAPTTPET-----PPPTTSEVSPNTTKPEPT 844
Db	629 PPTYKPKITYPPTYK-PKPSYPPSYRPKITYPPTYKPKKSYPOAKYKSGSYPPSYQPKT 687	Db	629 PPTYKPKITYPPTYK-PKPSYPPSYRPKITYPPTYKPKKSYPOAKYKSGSYPPSYQPKT 687
Qy	845 IHKSPDESTPELSAEPNTKALENSKPEGVPPTTKTPAATKPEMTTITAKDKT----TERDL 900	Qy	845 IHKSPDESTPELSAEPNTKALENSKPEGVPPTTKTPAATKPEMTTITAKDKT----TERDL 900
Db	688 Y---PPSKPKKKTYPPTYK-----PKISYPTPK-----TKPSYASVKRKTSYPTPKPK 735	Db	688 Y---PPSKPKKKTYPPTYK-----PKISYPTPK-----TKPSYASVKRKTSYPTPKPK 735
Qy	901 RTTPETTTAAPKMKTEAATTEKTTESKIIWATTTQVTSITTTQDITTPKIIITLKITLAPK 960	Qy	901 RTTPETTTAAPKMKTEAATTEKTTESKIIWATTTQVTSITTTQDITTPKIIITLKITLAPK 960
Db	736 ISYESTYKAKP-----GYPPPTYKPKPSYASSYKPKIRPYTPYKPKPSYASSYK-----PK 785	Db	736 ISYESTYKAKP-----GYPPPTYKPKPSYASSYKPKIRPYTPYKPKPSYASSYK-----PK 785

QY	961	VTTKTKTITTTTEIMNKPEETAFAKPKDRATNSKAITPKQ	---KPTKAPKKP-TSPKPKP-T	1015
Db	786	I-----RYPPTYKKP	---SYASSYKKIRYPPTYKKPKPSYASSYKKP	827
QY	1016	MPRVKPKTT--PTPRKWTSTMPELNPTSRIATAEAMLTQTRPNQ		1057
Db	828	YPPTYKKISYPPPTYKKKAITPPYKKPKAISYPPAYKKISVPSQ		871

RESULT	12
NFH_RAT	
ID	NFH_RAT
AC	P16884; Q63368; STANDARD; PRT; 831 AA.
DT	01-AUG-1990 (Rel. 15, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE	(NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
GN	NEFH OR NFH.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RP	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RC	MEDLINE=89065087; PubMed=3143506;
RA	Brann K.C., Robinson P.A., Wilson D., Anderson B.H.;
RA	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT	Identification of putative phosphorylation sites.";
RT	FEBS Lett. 241:213-218(1988).
RL	[2]
RN	SEQUENCE OF 37-831 FROM N.A.
RP	MEDLINE=88309090; PubMed=2457365;
RX	Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,
RA	Jolles P.;
RA	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT	in situ detection.";
RT	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
RL	[3]
RN	SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RP	MEDLINE=87080760; PubMed=2878828;
RX	Robinson P.A., Wilson D., Anderson B.H.;
RA	"Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT	(NF-H).";
RT	FEBS Lett. 209:203-205(1986).
RL	[4]
RN	SEQUENCE OF 318-831 FROM N.A.
RP	MEDLINE=89184647; PubMed=2928342;
RX	Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA	Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RA	"Cloning of a cDNA encoding the rat high molecular weight
RT	neurofilament peptide (NF-H): developmental and tissue expression in
RT	the rat, and mapping of its human homologue to chromosomes 1 and
RT	22.";
RT	Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
RL	[5]
RN	FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC	NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC	SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC	PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC	THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC	OF AXONAL CALIBER.
CC	PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC	OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC	LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC	COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC	SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC	CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
CC	ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).

- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL OF CA+2 PER MOL OF PROTEIN.

- SUBUNIT: HOMODIMER (PROBABLE).

- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.

- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF COMPOUND EYES AND OCELLI.

- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL DEVELOPMENT.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; L02111; AAA28405.1; -;  
EMBL; L05080; AAA28420.1; -;  
PIR; A47282; A47282.  
FlyBase; FBgn0010218; Cpn.  
Calcium-binding.  
KW CONFLICT 36 36 A -> AVAPAVA (IN REF. 2).  
FT CONFLICT 43 43 I -> T (IN REF. 2).  
FT CONFLICT 64 64 I -> V (IN REF. 2).  
FT CONFLICT 76 76 P -> A (IN REF. 2).  
FT CONFLICT 100 100 T -> P (IN REF. 2).  
FT CONFLICT 126 127 VQ -> AP (IN REF. 2).  
FT CONFLICT 154 154 I -> V (IN REF. 2).  
FT CONFLICT 160 160 S -> T (IN REF. 2).  
FT CONFLICT 534 534 A -> E (IN REF. 2).  
FT CONFLICT 699 699 I -> T (IN REF. 2).  
FT CONFLICT 703 703 V -> L (IN REF. 2).  
FT CONFLICT 721 721 D -> E (IN REF. 2).  
SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;

Query Match 7.2%; Score 530; DB 1; Length 865;  
Best Local Similarity 26.2%; Pred. No. 2.5e-14;  
Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

QY 298 APTSKVLAKP-TPK---AETTTKGPALTTPKEPTTPKEPASTTPKEPTTPKISAPT 353  
Db 8 SPVSAPVAAPVTSVAAPVQVVSAPVAP-----APAAPVAVTPVAPPTLASVQPAT 61  
QY 354 PKEPAPTTKSAPTP-KEPAPTTTKEPAPTPKEPAPTTTKEPAPT-----TTKSAPTTP 408  
Db 62 VTIPAPAPIAAASVTPVASVAPVVAAPT-----FAASPVSTPVAVAPVAVAPVAP 116  
QY 409 KEPAPTPKKPAP-----TTKEPAPTPKEPT-----TTKEPAPTTKEPAPT 454  
Db 117 PVAATTPVQVIAAPVATPPVAAASAPTAAPVTPVISPIASPPVVPANTTPVAPV 176  
QY 455 PKEPAP-PTAPKKPAPTPKEPAPTPKEPAPTTTKEP-PSPTTPKEPAP-----TTKSA 506  
Db 177 AAVPAAPVVPVLAAPVAPVAPVVAETPAPPPVAIEIVATIECPVAPLIEVSVVATK 236  
QY 507 PTTTKEP---APTTTSAPTTKEPSPTTKEPA-PTTKEPAPTPKKPAPTTKEPAP 562  
Db 237 PLAAAEPPVVAAPATTPVVAAPAAASPHVSAPVAVETAVVAPVSASTEPVAAALTTAP 296  
QY 563 TTP-----KEPAPTTTCKKAPAPKEP---APTTPKET-----APTTPKKLPT 603  
Db 297 ETAPALPVVAESQVAVANTVATPTTAPETPTAPVVAETPEVASVAVAEETTPVWPV 356  
QY 604 TPEKL-APTTPKEPAPTTPEELAPTTPEPTTTPTEEP---APTTPKAAAPNTKEPAPT 659  
Db 357 AAESIPAPVATTPVAT---LAVTDPDVVASVAPVAVELPVVAPSPVSAVATPVDLAPP 413  
QY 660 TKEPAPTTTPKEPAPTTTPKETAPTTP-KGTAPTTLKE-----PAPTTTPKKPAPKE 708

Db 414 V----LPPVAAEPVPAVVAEETPETPAPASPVITIAALDIPEVAPVIAAPSDAPAEPSA 469  
QY 709 LAPTTTKEPTSTSDKAPATTPKGTAPTTP-----KEPAPTTTPKEP---APTTPKGTAPT 760  
Db 470 AAPIVSTTP--TTASVPETTPAPPAVPTPIDVSLSEAAIETPVAPPVEVTTTEVAVADV 527  
QY 761 TLKEPA-----PTTPKKKAPKELAPTTT-----KGPTSTSTSDKAPATT-----PKETAPT 806  
Db 528 APPEAAADLIIEPVEPPAPIDLLLEQTTSVPAVEAAESTSS--PIPETSLPPNEAVASP 585  
QY 807 PKEPAPTTPKKAPPTPETPPPTSEVSTPTTKEPTTHKSPD-----ESTPEL 856  
Db 586 EVAVAPITAEPIP-----EPEPSLATPTPEIPVEAPVIVQEAADAVAVPVTETSTIPET 641  
QY 857 SAEPTPKALNSKPEKGVPTTKTAPATKPEMTTAKDKTKTERDLRTP--ETTAAAPKMT 914  
Db 642 TVE-FPEAVAEKVLDPAL--TEAPVTTQEPDVANINDGAPATEI--TTPAVEIVTAAAEVS 697  
QY 915 KETATTTTEKTESKITATTQVTTST-----TTQDTPPKITTLKTTTLA-PKVTTTKKT 967  
Db 698 DIAIPVIDPPVPOEIAVAEIPETDTKPAEIVIEQSTIPEAPVPEVSKYAEPVISEAPAA 757  
QY 968 ITTTEIMNKPEETAKPKDRATNSKATTPKPKPTKAPKPTSTKKPKTMPVRVKPKPTPT 1027  
Db 758 EVPITAGDNDNTSGVISEVPTIAEKPVVEEPTS--EIEQSSSPS--DSVPVAKITPL 813  
QY 1028 PRKMTSTMPELNPTSRIAEAM---LQTTTRPNQTPNSKLVEVN-----PKS 1070  
Db 814 LRDLQTTDVSLLAIAATLDAIGEKLKDKARKQVQVMDRLCEIEKILGPPKS 864

RESULT 11  
FPI\_MITCO  
ID FPI\_MITCO STANDARD; PRT; 872 AA.  
AC Q25434;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCFP1).  
GN FPI.  
OS Mytilus coruscus (Sea mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
OC Mytiloidae; Mytilidae; Mytilus.  
OX NCBI\_TaxID=42192;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Foot;  
RX MEDLINE=96394686; PubMed=8798340;  
RA Inoue K., Takeuchi Y., Takeyama S., Yamaha E., Yamazaki F., Odo S.,  
RA Harayama S.;  
RT "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and  
its evolutionary implications";  
RL J. Mol. Evol. 43:348-356(1996).  
CC - FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
CC - SUBCELLULAR LOCATION: SECRETED.  
CC - TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
CC - DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
CC - PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND  
ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLALANINE).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D63777; BAA09850.1; -;  
InterPro; IPR002964; Adhesive\_plaq.

01-NOV-1991 (Rel. 20, Last sequence update)  
01-NOV-1991 (Rel. 20, Last sequence update)  
20-AUG-2001 (Rel. 40, Last annotation update)  
STALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).  
TCNA.  
Trypanosoma cruzi.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
NCBI\_TaxID=5693;  
[1]  
SEQUENCE FROM N.A.  
STRAIN-SILVIO X-10/4;  
MEDLINE=9127609; PubMed=1711561;  
Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D., Prioli R.P.;  
"The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neuraminidases, YMD repeats of the low density lipoprotein receptor, and type III modules of fibronectin.";  
J. Exp. Med. 174:179-191(1991).  
[2]  
SUBCELLULAR LOCATION.  
MEDLINE=91376547; PubMed=1896773;  
Prioli R.P., Mejia J.S., Aji T., Akawa M., Pereira M.E.A.;  
"Trypanosoma cruzi: localization of neuraminidase on the surface of trypomastigotes.";  
Trop. Med. Parasitol. 42:146-150(1991).  
-!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN PARASITE INVASION OF CELLS.  
-!- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLEURAMINYL RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS, GLYCOLIPIDS OR COLOMINIC ACID.  
-!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POSSIBLE).  
-!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.  
-!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE INCA PROTEIN.  
-!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.  
-!- SIMILARITY: CONTAINS 3 BNR REPEATS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M61732; AAA30255.1; -;  
PIR; JH0557; JH0557.  
HSP; P29768; IDIL.  
InterPro: IPR002860; BNR.  
Pfam: PF02012; BNR; 2.  
Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;  
Phosphorylation.  
DOMAIN 1 457  
REPEAT 23 34  
REPEAT 163 174  
REPEAT 209 220  
REPEAT 458 588  
DOMAIN 589 1120  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
CARBOHYD 342 342  
CARBOHYD 394 394  
CARBOHYD 1125 1125  
SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;  
-----  
Query Match 7.2%; Score 530.5; DB 1; Length 1162;  
Best Local Similarity 29.4%; Pred. No. 3e-14;  
Matches 177; Conservative 100; Mismatches 212; Indels 113; Gaps 34;  
-----

Query Match	7.2%;	Score 530.5;	DB 1;	Length 1162;
Best Local Similarity	29.4%;	Pred. No. 3e-14;		
Matches 177;	Conservative 100;	Mismatches 212;	Indels 113;	Gaps 34;

Gene 58:307-314(1988).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=89089136; PubMed=3145094;  
RX Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;  
RA "the structure of the largest murine neurofilament protein (NF-H) as  
RT revealed by cDNA and genomic sequences.";  
RL Brain Res. 464:217-231(1988).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=SWISS WEBSTER; TISSUE=Brain;  
RA Carden M.J.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS  
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M24496; AAA39813.1; -  
DR EMBL; M23349; AAA39813.1; JOINED.  
DR EMBL; M24494; AAA39813.1; JOINED.  
DR EMBL; M24495; AAA39813.1; JOINED.  
DR EMBL; M35131; AAA39808.1; ALT\_FRAME.  
DR EMBL; Z31012; CAA83229.1; -  
DR PIR; J03368; QFM5H.  
DR PIR; A43778; A43778.  
DR MGD; MGI:97309; Nfh.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
KW Repeat.  
FT DOMAIN 1 97 HEAD.  
FT FT DOMAIN 98 408 ROD.  
FT FT DOMAIN 409 1087 TAIL.  
FT FT DOMAIN 436 517 GLU-RICH (ACIDIC).  
FT FT DOMAIN 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.  
FT FT DOMAIN 887 1087 GLU/LYS-RICH.  
FT FT DOMAIN 98 129 COIL 1A.  
FT FT DOMAIN 130 141 LINKER 1.  
FT FT DOMAIN 142 239 COIL 1B.  
FT FT DOMAIN 240 261 LINKER 12.  
FT FT DOMAIN 262 283 COIL 2A.  
FT FT DOMAIN 284 287 LINKER 2.  
FT FT DOMAIN 288 408 COIL 2B.  
FT FT CONFLICT 133 133 K -> QA (IN REF. 2 AND 3).  
FT FT CONFLICT 199 199 A -> AR (IN REF. 2 AND 3).  
FT FT CONFLICT 281 281 S -> T (IN REF. 2 AND 3).  
FT FT CONFLICT 492 492 L -> G (IN REF. 2 AND 3).  
FT FT CONFLICT 551 551 P -> PREAKSP (IN REF. 3).  
FT FT CONFLICT 689 712 MISSING (IN REF. 3).  
FT FT CONFLICT 714 714 G -> A (IN REF. 3).  
FT FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).  
FT FT CONFLICT 814 814



```

OY 913 MKETATTEKTESKITATTTQVSTTODTTPKITTLLKPTTLAPKVTTLTKTTTTE 972
DB 748 -TYKAKPTYPPTKAK-----PSYPTKPKPSYPTKSKSYPSYKPKKTYPT- 798
OY 973 INKKEETAKPKDRATNSKATTPKQPKPKAKK---PTSTKPKPTMPRVKPKKTPPP- 1028
DB 799 --YKFKLTPPYK-----PKSPSYKPKKITYPTKYKLPKSPYPTKSKTSKTSPT 848
OY 1029 -----RKMTSTPELNPTSR 1043
DB 849 YNKISYPSYKAKTSYPAKPKTR 874

RESULT 6
ZAN_HUMAN
ID ZAN_HUMAN STANDARD: PRT; 2700 AA.
AC O9Y493; 000218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZONADHESIN (FRAGMENT).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE=99018118; PubMed=9799793;
RA Glocker G., Scherer S., Schattevoy R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes."
RL Genome Res. 8:1060-1073(1998).
RN [2]
RP SEQUENCE OF 2338-2700 FROM N.A.
RX TISSUE-Testis;
RA MEDLINE=97271566; PubMed=9126492;
RA Geo Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLE FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: IN TESTIS, PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOIA OR PROMOTING ADHESION TO THE OVUDUCTAL ISTHMS.
CC -1- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC -1- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF053356; AAC78790.1; -
DR EMBL; U83191; AAC51208.1; -
DR MIM; 602372; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000998; MAM.

```

```

DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR002919; TIL.
DR InterPro: IPR003328; TILA.
DR InterPro: IPR001007; VMFC.
DR InterPro: IPR001846; Vwd.
DR Pfam: PF00629; MAM; 3.
DR Pfam: PF01826; TIL; 4.
DR Pfam: PF02345; TILA; 4.
DR Pfam: PF00094; Vwd; 4.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00137; MAM; 2.
DR SMART: SM00014; VMC; 1.
DR SMART: SM00016; VMC_def; 3.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS50060; MAM_2; 4.
KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT NON_TER 1 1
FT DOMAIN <1 109 MAM 1.
FT DOMAIN 112 136 MAM 2 (PARTIAL).
FT DOMAIN 161 326 MAM 3.
FT DOMAIN 322 446 MAM 4.
FT DOMAIN 483 951
FT DOMAIN 953 1065
FT DOMAIN 1066 1454
FT DOMAIN 1455 1861
FT DOMAIN 1862 2292
FT DOMAIN 2293 2684
FT DOMAIN ? 74
FT CARBOHYD 74 74 EGF-LIKE.
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2374 2379 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 2700 2700
SQ SEQUENCE 2700 AA; 293013 MW; 80E60CC0B12277B1 CRC64;

Query Match 7.5%; Score 555.5; DB 1; Length 2700;
Best Local Similarity 32.6%; Pred. No. 6,4e-15;
Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;

OY 233 PINPRSLPPNSDTSKESLVNKKETLYEKKTTTNNKQTSIDGKETSAAKETQSIK 292
DB 445 PKVLPDLPPVSVSS-----TGPSETTGTEPNTSTK-----KPTVSIK 487
OY 293 SAKDLAPTSKVLAKPTPAETTTKGPALTTPKEPTTPPKKEPASTTPKEPT-----PTTI 347
DB 488 SYTEKPT-----YKKEKPTIPEKPTISTEK-----PTIPSEKPNPNSKPTIIPSKPTIL 539
OY 348 KCAPTPKEPAPATTTKSAPTTTKEPAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYK 406
DB 540 TEKPTIPE-KPTIPEKPTISTEKPTVTEP--PTTPEETTYMEBPVTEKPT 596
OY 407 TPKEPAPTPPKKAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYK 466
DB 597 --EKRSITPEK---PTIMEETIISTEKPTICPEKPTIPEK---PTIPEKSTISPEK- 647
OY 467 APPTKEPAPTPPKKAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYK 526
DB 648 -PTTPE-KPTIPEKPTISTEKPTIPEK-PTISPEKLTITPEKLTITPT--EKPTIPT 701
OY 527 EPSPTTKEPAPTPPKKAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYKKEPAPATTYK 581
DB 702 EKPTISTEE-PTTPEETTYISTEKPSIPEKPTIPEETTYISVEETTYISTEKLTIPM-- 757

```

```

OY      888 TTANKDKTERDOLRTPEPT-----TAAPKMKKEATITTEKTESKIATTOYTST 940
          | :   | :   | :   | :   | :   ||| :|| 
Db       754 --PAHGVTASADTIRPAGGTAPPRAHGVTSAP---DIRPARGSTAPPAHGVTSAPIPTRA 807
          | :   | :   | :   | :   | :   ||| :|| 
OY      941 TODPP--FKTTTLTKTTTLAKPVTTTKKITLT----ETMNPDEEFAEKDRATNSKATTPK 966
          | :   | :   | :   | :   | :   ||| :|| 
Db       808 PGSTARPHAGVTSAPDIIRPARGSTAPPRAHGVTSAPDIIRPARGSTAPPRAHGVTSAPIPTRPA 867
          | :   | :   | :   | :   | :   ||| :|| 
OY      997 POKRTPARKRPSTKKRKPTMRVRKRKPCTPPPKNMTSMELPNSTKLAEAMLOTTTRPN 1056
          | :   | :   | :   | :   | :   ||| :|| 
Db       868 P--GSTAPPAGGVTSAPDPTR--APGSTAPPAGGVTSAPDPTRA-----PG 909
          | :   | :   | :   | :   | :   ||| :|| 
OY      1057 QT-PMSKIVENVPSESDGAEGEPNHLILRHVFMEFVEPDMDYLPRVDNOGIINPMI 1115
          | :   | :   | :   | :   | :   ||| :|| 
Db       910 STAPPAGHVTSAPDIIRPARGSTAPPRAH-----GVISAPDNRRALSGSTA-----PV 955
          | :   | :   | :   | :   | :   ||| :|| 
OY      1116 SDEFNICNGKPYDGILTLLRNGLTVAFRGHYFWMLSPESPSS 1156
          | :   | :   | :   | :   | :   ||| :|| 
Db       956 HNVTSASGSASGSASTLVHNCHSARATTPPAKSSTPFPSIPS 996
          | :   | :   | :   | :   | :   ||| :|| 

RESULT    5
FPL_MYTEd FPL_MYTEd STANDARD; PRt; 875 AA.
AC         025460;
DT        30-MAY-2000 (Rel. 39, Created)
DT        30-MAY-2000 (Rel. 39, Last sequence update)
DT        20-AUG-2001 (Rel. 40, last annotation update)
DE        ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
GN        PROTEIN 1) (MEFP1) (FRAGMENT).
OS        FPL.
OC        Mytilus edulis (Blue mussel).
OX        Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
           NCBI_TaxID=6550;
RN        [1]
RX        SEQUENCE FROM N.A., MEDLINE=91025829; PubMed=1367451; Strausberg R.L.;
RT        Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
RL        "Structural and functional repetition in a marine mussel adhesive
           protein." ;
RM        Biotechnol. Prog. 6:171-177(1990).
[2]
PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
MEDLINE=83135732; PubMed=6298211;
RX        Waite J.H.;
RA        "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
           hydroxyproline-containing decapeptide in the adhesive protein of the
           mussel, Mytilus edulis L.";
RJ        J. Biol. Chem 258:2911-2915(1983).
CC        -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
           PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
           ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
           FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
           SUBCELLULAR LOCATION: SECRETED.
           TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
           DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
           PTM: THE DECAPEPTIDE A-K-P-S-Y-P-T-X-R IS POST-TRANSLATIONALLY
           MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
           HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYROXYPHENYLALANINE
           (DOXA) DERIVED FROM TYROSINE.
```

[illegible]



RP SEQUENCE OF 1-89 FROM N.A.  
RC TISSUE-Lung;  
RX MEDLINE=96181716; PubMed=8604237;  
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,  
RA Lee L.N., Luh K.T., Wu C.W.;  
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and  
RT tissues";  
RL Oncology 53:118-126(1996).  
RN (12)  
RP SEQUENCE OF 1-46 FROM N.A.  
RC TISSUE-Breast carcinoma;  
RA Bulveira L., Liu O., Lugmani Y.A., Gomm J.J., Coombes R.C.;  
RL Submitted (Oct-1992) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CYTOSKELETON.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM  
IS ALSO PRODUCED.  
CC -1- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE  
SPLICING.  
CC -1- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL  
TUMORS, SUCH AS BREAST CANCER.  
CC -1- PTM: HIGHLY GLYCOSYLATED (N-AND O-LINKED CARBOHYDRATES AND SIALIC  
ACID).  
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC. IT  
VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE  
MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation-  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; J05582; AAA60019.1; -;  
DR EMBL; M32738; AAA35804.1; -;  
DR EMBL; M32739; AAA35806.1; -;  
DR EMBL; J05581; AAA59876.1; -;  
DR EMBL; M61170; AAB53150.1; -;  
DR EMBL; X52229; CA36478.1; ALT\_SEQ.  
DR EMBL; X52228; CA36477.1; ALT\_SEQ.  
DR EMBL; M35093; AAB59612.1; ALT\_SEQ.  
DR EMBL; Z17324; AAB78972.1; -;  
DR EMBL; Z17325; CA278973.1; -;  
DR EMBL; M31823; AAA35757.1; -;  
DR EMBL; S81781; AAD14376.1; ALT\_INIT.  
DR EMBL; S81736; AAD14369.1; ALT\_INIT.  
DR EMBL; M21868; AAA59874.1; ALT\_SEQ.  
DR PIR; A35175; A35175.  
DR PIR; B35175; B35175.  
DR PIR; S10218; S10218.  
DR GLYCOSULEDB; P15941; -;  
DR MIM; 158340; -;  
DR MIM; 113720; -;  
DR InterPro; IPR000082; SEA.  
DR Pfam; PF01390; SEA; 1.  
DR SMART; SM00200; SEA; 1.  
DR PROSITE; PSS0024; SEA; 1.  
KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
KW Repeat; Alternative splicing.  
FT SIGNAL 1 23  
FT CHAIN 24 1255  
FT DOMAIN 24 1162  
FT TRANSLEM 1163 1186  
FT DOMAIN 1187 1235  
FT DOMAIN 81 960  
FT DOMAIN 1034 1151  
FT CARBOHYD 957 957  
FT CARBOHYD 975 975  
FT CARBOHYD 1029 1029  
FT CARBOHYD 1055 1055

FT CARBOHYD 1133 1133  
FT VARSPLIC 19 19  
FT VARSPLIC 20 22  
FT VARSPLIC 20 31  
FT VARSPLIC 126 905  
FT VARSPLIC 1077 1087  
FT VARSPLIC 1088 1255  
FT CONFLICT 2 2  
FT CONFLICT 134 134  
FT CONFLICT 154 154  
FT CONFLICT 1021 1021  
FT CONFLICT 1251 1251  
SQ SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;  
Query Match 8.8%; Score 651; DB 1: Length 1255;  
Best Local Similarity 27.8%; Pred No. 7.1e-19;  
Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;  
QY 250 TSLVNETTETVETKETTNNKQTSIDGKEKTSANETOSIETKSARD-LAPTSKYLAPPT 308  
DB 16 TLVTV-----VTGSGHASTPGEGKEKTSATORSSVPSSTKKNVMSVSLSSHS 65  
QY 309 P-KAETTKGP--ALTTKEP-----TPPTKPEASTTTKEPPTTT--K 348  
DB 66 PGGSSSTTGQGDVTLAPATEPAGSSAATWGQDVTSVPYTRPALGSTTPAHVDIAPDNK 125  
QY 349 SAPPTKPEAPPTTKSAPPTKPEAPPTTKP-----APPTKPEAPPTTKPEAPPTTK 402  
DB 126 PAFGSTAPPAHGVY-SAPDT--RPAAGSTAPPAHGVTSAPDT--RPAAGSTAPPAHGVY- 179  
QY 403 SAPPTKPEAPPTTKP-----KKAAPTTPKEP-----APTTKPEPTPTT- 440  
DB 180 SAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVY 239  
QY 441 ----KEPAPTTKEP-----APTTKPEAPPTTKP-----KKAAPTTPKEPAPPTTKP--E 481  
DB 240 SAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVY 299  
QY 482 PAPPTTKPEPTPTTKPEAPPTTKSAPTTKPEAPPTTKSAPPTTKPEPTPTTKPEPTTKPE 536  
DB 300 SAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVY-SAPDT--RPAAGSTAPPAH 356  
QY 537 ----APTTKPEAPPTTKP-----KKAAPTTPKEPAPPTTKP-----EAPPTTKKAPPAPEKPEP 584  
DB 357 GVTSAAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAH 416  
QY 585 APPTTKPEAPPTTKPKLPTTPEKLAPTTPEKAPPTTPEELAPTTPEPTPTTPEEPAPPTT 644  
DB 417 GVTSAAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAP--PAHGVTSAPDTRPAAGSTA 472  
QY 645 PKA-----AAPTTPKEPAPPTTKP-----EAPPTTKPEAPPTTKPEAPPTTKGTAAPTTLKEP 696  
DB 473 PPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAP--PAHGVTSAPDTRPAAP 528  
QY 697 APTTP-----KKPAPELAPTTTKEPTSTISQKAP--TTPKGTAPPTTKEPAPT 744  
DB 529 GSTAPPAHGVTSAPDTRPAAGSTAP--PAHGVTSAPDTRPAAGSTAPPAHGVTS---APD 583  
QY 745 TPKEPAPPTTKG-----TAPPTTLKEPAPPTT-----KKPAPELAPTTTGTGPTSTT 790  
DB 584 TRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAP--PAHGVTSAP 642  
QY 791 SDRAP--TTPK-----ETAPPTTKEPAPPTT-----KKPAPELAPTTTGTGPTSTT 834  
DB 643 DTRAPAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAPDTRPAAGS--TAPPAHGVTS 700  
QY 835 TPTTKEPTT-----IHSPDESTPELSAEPKALENSPEKPGVPTTKTAATPEM 887  
DB 701 APDTRPAAGSTAPPAHGVTSAPDTRPAAGSTAPPAHGVTSAP-----DTRAPAGSTAP-- 753

```

Db      612 PPSSTSTSSAPVT-----SSTSSAP-VTPSSSTSSAPVPPSSSTSSA 666
Qy      640 PAP-----TTPAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTL 693
Db      667 PVPPTSSSTSSAPVT-----SSTSSAPVT-----SSTSSAPVTPSSSTSS 718
Qy      694 KEPAPTTPKPAKELAPTTTKEPTSTSDKPAPTTPKTAPTTPKEPAP--TTPKEPA 750
Db      719 SAVPTPSSSTSSAPVTPPS--SSTSSAPVTSSTSSAPVTPSSSTSSA 777
Qy      751 PT-TPKGTAPTTTKEPAPTTPKPAKELAPTTTGTPTSTGTPAPTPKTAPTTPKE 809
Db      778 PVPPTSSSTSSAPVPPSSSTSSAPVTPSSSTSSAPVTPSSSTSSA 837
Qy      810 PAP-----TTPKPAPTTPPTTSEVSTPT-----TTPKEPTTHKSPDESPELSAE 859
Db      838 PVPPTSSSTSSAPVSSSTSSAPVTPSSSNTSSAPSI--PFSSSTESFT 894
Qy      860 PPKALENSPEKPEVPTTKTPAATKPBMTTAKDKTTERDLRTTPTTTPAATKETA 919
Db      895 GT-TTPSSSKYPGQTESEVSTETTIVPTKTTTSVTPSTTTTITVCSGTINSAGE 953
Qy      920 TTPKTESKTA--TQVSTSTODTTPKITTTLKTTLAPVTT--TKKTTT--EI 973
Db      954 TTSKSPKVTIVPTTTSVTSSTTTTITVCSGTINSAGETTSKSPKTTTTPC 1013
Qy      974 MKPPEETAKPKDRATNSKATTPKPKP-----TKAPKPTSTKPKPTM 1016
Db      1014 STSPSEEA-----SESTTSPTEPTVTVSTVTVTEYSTKPGETTTTFTVKNI 1065
Qy      1017 PVRKPKPTTPPKRMSTMEPLNTSRIAMOTTRPMOTPSKLEVNPKSEDA 1076
Db      1066 PTTYLTATPTP--SVTVTNTPTT--ITTVCSGT-----NSAGETSGC 1109
Qy      1077 EGEPMHLLPHEMPEVTDMVLPVNOGIIINPLSDETMICGKPEVDGTLRLNG 1136
Db      1110 SPKVTITVCSGTGTETATLTVAATVTVTSSGNTSA-GKTTTGTTSV 1168
Qy      1137 TLVAFRGHYFWMLSFSPSPARRITEWGISPIDVTFRC--NCEGT 1184
Db      1169 TT-----YVTLTAPSAVTPATN-----AVPTTITT--TECSAATNAAGET 1207

RESULT 4
ID      MUC1_HUMAN          STANDARD;          PRT; 1255 AA.
AC      P15941; P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4J2;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DE      20-APR-2001 (Rel. 40, last annotation update)
DE      MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE      (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE      (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
DE      REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE      DF3).
GN      MUC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Pancreas;
RX      MEDLINE=90368716; PubMed=2394722;
RA      Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT      "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL      J. Biol. Chem. 265:15294-15299(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90202794; PubMed=2318825;
RT      "Epistatin, a carcinoma-associated mucin, is generated by a

```

```

RT      polymorphic gene encoding splice variants with alternative amino
RT      termini.";
RL      J. Biol. Chem. 265:5573-5578(1990).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      MEDLINE=90368715; PubMed=1697589;
RA      Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA      Peat N., Burchell J., Pemberton L., Ialenti E.-N., Wilson D.;
RT      "Molecular cloning and expression of human tumor-associated
RT      polymorphic epithelial mucin.";
RL      J. Biol. Chem. 265:15286-15293(1990).
RN      [4]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91097524; PubMed=2268309;
RA      Lancaster C.A., Peat N., Duhig T., Wilson D.;
RA      Taylor-Papadimitriou J., Gendler S.J.;
RT      "Structure and expression of the human polymorphic epithelial mucin
RT      gene: an expressed VNTR unit.";
RL      Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN      [5]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      MEDLINE=90276413; PubMed=2351132;
RA      Wreschner D.H., Hareven M., Tsarfaty I., Smorodinsky N.,
RA      Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA      Keydar I.;
RT      "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT      may generate multiple protein forms.";
RL      Eur. J. Biochem. 189:463-473(1990).
RN      [6]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      MEDLINE=90276414; PubMed=2112450;
RA      Hareven M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA      Zilhan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT      "A transcribed gene, containing a variable number of tandem repeats,
RT      codes for a human epithelial tumor antigen. cDNA cloning, expression
RT      of the transfected gene and over-expression in breast cancer
RT      tissue.";
RL      Eur. J. Biochem. 189:475-486(1990).
RN      [7]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9103045; PubMed=1688329;
RA      Tsarfaty I., Hareven M., Horev J., Zaretsky J., Weiss M.,
RA      Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT      "Isolation and characterization of an expressed hypervariable gene
RT      coding for a breast-cancer-associated antigen.";
RL      Gene 93:313-318(1990).
RN      [8]
RP      PARTIAL SEQUENCE FROM N.A.
RX      MEDLINE=88330762; PubMed=3417635;
RA      Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA      Burchell J.;
RT      "A highly immunogenic region of a human polymorphic epithelial mucin
RT      expressed by carcinomas is made up of tandem repeats.";
RL      J. Biol. Chem. 263:12820-12823(1988).
RN      [9]
RP      SEQUENCE OF 1-169 FROM N.A.
RX      MEDLINE=90088473; PubMed=2597151;
RA      Abe M., Siddiqui J., Kuze D.;
RT      "Sequence analysis of the 5' region of the human DF3 breast
RT      carcinoma-associated antigen gene.";
RL      Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN      [10]
RP      SEQUENCE OF 1-109 FROM N.A.
RC      TISSUE=Thyroid;
RX      MEDLINE=96183746; PubMed=8608966;
RA      Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT      "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT      transcriptase polymerase chain reaction of the Muc1 gene.";
RL      Int. J. Cancer 66:55-59(1996).
RN      [11]

```

CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC	GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC	MIR RELEASE OF BETA-D-GLUCOSE.
CC	-1- SIMILARITY: TO S.POMBE SPCC215.138.
CC	-1- SIMILARITY: SOME, TO S.POMBE SPCC215.13C.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; Z38061; CAA86176.1; -
DR	EMBL; M16164; AAA35014.1; -
DR	EMBL; M16165; AAA35015.1; -
DR	EMBL; X13857; CAA32069.1; -
DR	PIR; B26877; B26877;
DR	PIR; A26877; A26877;
DR	PIR; S48478; S48478.
DR	SGD; S0001458; MUC1.
KW	Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
FT	Signal: Multigene family.
FT	SIGNAL
FT	CHAIN
FT	POTENTIAL.
FT	DOMAIN
FT	CARBOHYD
FT	CARBOHYD
SO	SEQUENCE

1 21  
22 1367  
210 1367  
817 817  
874 874  
136110 AA; 136110 MW; 91C00DE2DBED61AA9D CNC64;

```

Query Match: 10.9% Score 809; DB 1; Length 1367;
Best Local Similarity 27.5%; Pred No. 6,1e-25;
Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53.

QY 61 DCDACQKKTKDK---CCPDYSEFCAEVYHNPTSPSSAKAPPPGASQSTIKSTRKSPKN 117
Db 174 DLSTGCNNYNDNOGHQSOTDPGFYWNIDCCNNCGCTKSTTSSTSS--STI----- 223
QY 118 KKKTKKVEISEETTEHSHVSENOESSSSSSSSSTTIKIKSSKNSAANRELQKKLYK 177
Db 224 ----TSTSSSSSTTSSSSSTTSSSSSTTSSSSSTT-----AAATPTTSCTEKPP 274
QY 178 DNKKNKTKKPPKPPVYVDAGSGLDNGBFKYTPPTDSTYGHNVKSTSPKITAKPINR 237
Db 275 PTTTSCTEKPP--DPHND-----TTPCT-----KKTTTTSKITKKTKTTPV 314
QY 238 PSLPNSDPSKSTSLTVNKETVEKETTNTTKQSTQCKEKTTSAKETOSIEKTSANDL 297
Db 315 PT--PSSSTTESS-----APVPPSSSTT-----ESSAPVSSSTESS--- 353
QY 298 APTSKVLAKPTPKPAETTTKGPALTTPKBPPTPTPKREPASTTPKEPTTIIASAPTTKREP 357
Db 354 APV-----PPSSSTTESSAPVSSSTTESSAPVSSSTTESSAPVPPSSSTTESS 407
QY 358 APTT-----TISAP-----TPKREPAITTKREPAITPKREPA---TTTKREPAITTKSA 404
Db 408 APVSTSTTESSAPVTSSTTESSAPVTS---TTESSAPVTSSTTESSAPVPPSS 463
QY 405 PTTKREPAITTPKKPAITPKREPAITTPKREPT--TPKREPAITPKREPAITTKREPAIT 461
Db 464 STTESSAPV-----SSTTESSAPV-----PTSSSTTESSAPV---SSTTESSAPV 512
QY 462 APKPAITPKREPAITPKREPAITTKKEPSTTPKREPAITTKSAPITTKREPAITTKSA 521
Db 513 -----PTSSSTTESSAPVPPS---SSTTESSAPVTS---TTESSAPVPPSS 559
QY 522 PTTKEPSPITTKREPAITPKREPAIT--TPKKPATITKREPAITPKREPAITTKKPAITA 580
Db 560 STSSSTTPVTS---TTESSAPVPPSSSTTESSAPVPT---PSSSTTESSAPA 611

```



CC This SMS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC -- send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

FT	CARBOHD	4339	4339	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4351	4351	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4362	4362	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4373	4373	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4422	4422	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4438	4438	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4502	4502	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4616	4616	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4627	4627	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4752	4752	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4787	4787	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4881	4881	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4888	4888	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4955	4955	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	4970	4970	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	5019	5019	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	5038	5038	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CARBOHD	5069	5069	N-LINKED	(GLCNAC. . .)	(POTENTIAL)
FT	CONFLICT	1351	1351	H -> L (IN REF. 3)		
FT	CONFLICT	1412	1412	T -> S (IN REF. 3)		
FT	CONFLICT	1449	1449	L -> P (IN REF. 3)		
FT	CONFLICT	1504	1504	M -> T (IN REF. 3)		
FT	CONFLICT	4192	4192	G -> S (IN REF. 2)		
SO	SEQUENCE	5179	AA: 540295	MM: 85CD1571FB9A56B3	CRC64:	

Query Match	16.0%;	Score 1188;	DB 1;	Length 5179;
Best Local Similarity	29.5%;	Pred. No. 4.2e-39;		
Matches 403;	Conservative 85;	Mismatches 536;	Indels 344;	Gaps 53;

[illegible]

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:03 ; Search time 47.4 Seconds  
(without alignments)  
1067.457 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_25\_1404

Perfect score: 7410  
Sequence: 1 QDLSSGAGCGEGYSRDATC.....ARATITRSGQTLSKWMYNC P 1380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

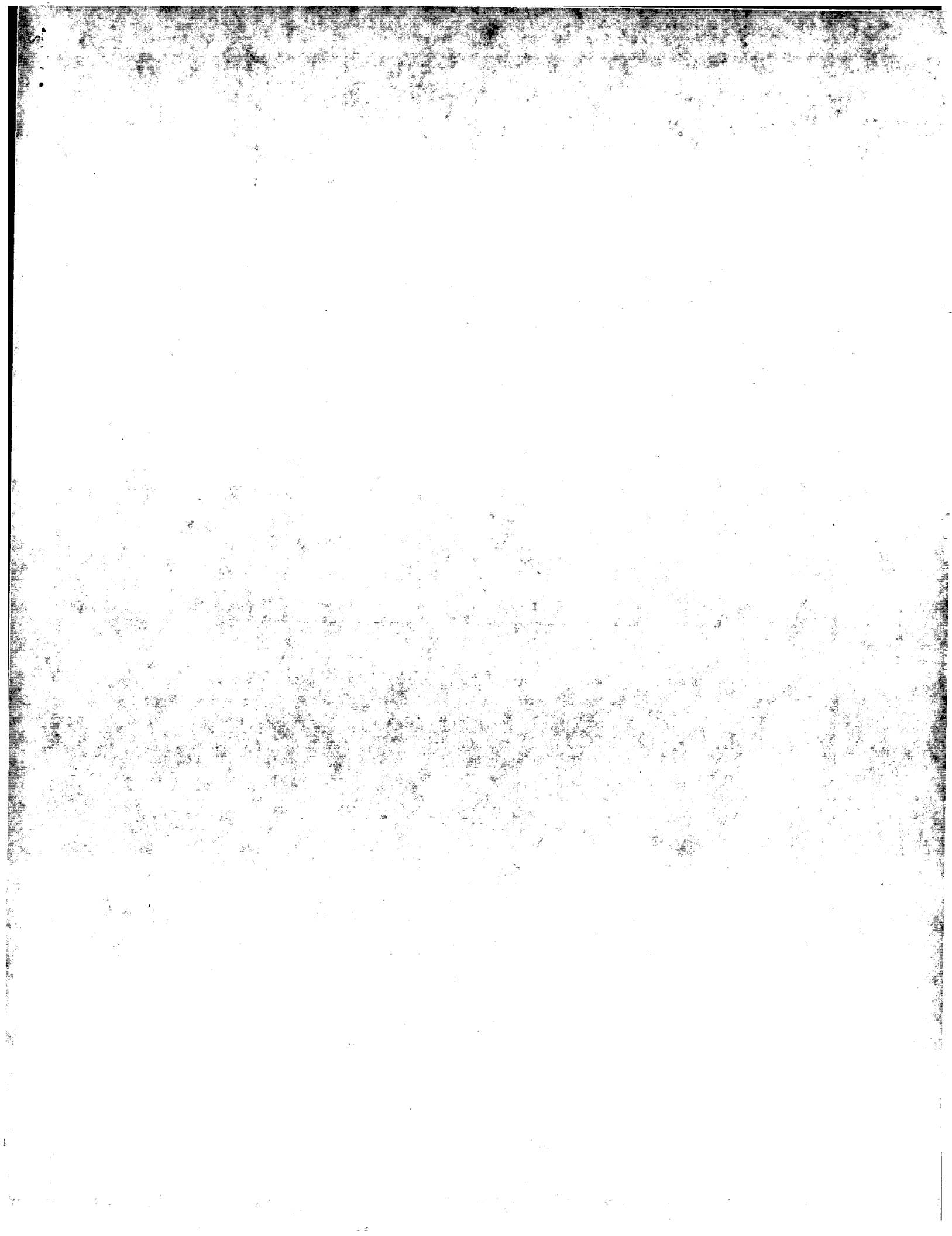
## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1188	16.0	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	12.8	1664	1 SIPL_CLOM	Q06852 clostridium
3	809	10.9	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	8.8	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	559	7.5	875	1 FPI_MYTED	O25460 mytilus edu
6	555.5	7.5	2700	1 ZAN_HUMAN	O9y493 homo sapien
7	551	7.4	620	1 EXTN_TOBAC	P13983 nicotiana t
8	530.5	7.2	1087	1 NFI_MOUSE	P19246 mus muscula
9	530.5	7.2	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.2	865	1 CPN_DROME	O02910 drosophila
11	518.5	7.0	872	1 FPI_MYTCO	O25434 mytilus cor
12	517.5	7.0	831	1 NFI_RAT	P16884 rattus norv
13	509.5	6.9	662	1 MUC1_XENLA	O05049 xenopus lae
14	503.5	6.8	1970	1 RPB1_HUMAN	P24928 homo sapien
15	502	6.8	1020	1 NFI_MOUSE	P12036 homo sapien
16	498.5	6.7	1970	1 RPB1_MOUSE	P08775 mus musculu
17	493.5	6.7	467	1 RPB1_CRICR	P14141 cricetus
18	490	6.6	2142	1 BAT2_HUMAN	P48634 homo sapien
19	488.5	6.6	826	1 SSP2_PLAYO	O01443 plasmodium
20	475.5	6.4	267	1 EXTN_MAIZE	P14918 zea mays (m
21	471.5	6.4	5376	1 ZAN_MOUSE	O08799 mus musculu
22	468.5	6.3	634	1 VRP1_CANAL	P37370 saccharomyc
23	467	6.3	817	1 VRP1_YEAST	P28968 equine herp
24	454	6.1	797	1 VGEX_HSEVB	P47179 saccharomyc
25	454	6.1	1161	1 YJ9P_YEAST	O10172 schizosacch
26	453.5	6.1	1794	1 YAVI_SCHPO	O00130 ictalurid h
27	448.5	6.0	670	1 VG50_HSV11	O092n3 homo sapien
28	446	6.0	1229	1 NI21_HUMAN	P34926 rattus norv
29	444.5	6.0	2774	1 MAPA_RAT	O27409 mytilus gal
30	442.5	6.0	751	1 FPI_MYTCO	O00268 homo sapien
31	439.5	5.9	1083	1 T2D3_HUMAN	O13428 homo sapien
32	436.5	5.9	1411	1 TCOR_HUMAN	P46013 homo sapien
33	434.5	5.9	3256	1 KI67_HUMAN	

34	433.5	5.9	3164	1 TEGU_HSV11	P10220 herpes simp
35	432.5	5.8	439	1 XP2_XENLA	P17437 xenopus lae
36	428	5.8	2517	1 NCR2_HUMAN	O9y618 h nuclear r
37	427.5	5.8	1251	1 YOUT_CAEEL	O09550 caenorhabdi
38	426	5.7	2715	1 TRX2_HUMAN	O9um66 homo sapien
39	425.5	5.7	1185	1 DRPL_HUMAN	P54259 homo sapien
40	424.5	5.7	1125	1 MAPA_MOUSE	P27546 mus musculu
41	424	5.7	307	1 SCS3_DROME	P02840 drosophila
42	423.5	5.7	3421	1 TEGU_HSEVB	P28955 equine herp
43	422.5	5.7	1183	1 DRPL_RAT	P54258 rattus norv
44	419.5	5.7	2476	1 ZAN_PIG	O28983 sus scrofa
45	411.5	5.6	907	1 VGP3_EBV	P03200 Epstein-Bar

## ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	MUC2_HUMAN	STANDARD; PRT; 5179 AA.
AC	Q02817; Q14878;	
DT	01-JUN-1994 (Rel. 29, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).	
GN	MUC2 OR SMUC.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Intestine;	
RX	MEDLINE=94132002; PubMed=8300571;	
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;	
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.	
RT	Identification of the amino terminus and overall sequence similarity	
RT	to prepro-von Willebrand factor."	
RL	J. Biol. Chem. 269:2440-2446(1994).	
RN	[2]	
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.	
RC	TISSUE=Colon;	
RX	MEDLINE=93016075; PubMed=1400449;	
RA	Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,	
RA	Kim Y.S.;	
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located	
RT	both upstream and downstream of its central repetitive region."	
RL	J. Biol. Chem. 267:21375-21383(1992).	
RN	[3]	
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.	
RX	MEDLINE=91358717; PubMed=1885763;	
RA	Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,	
RA	Petersen G.M., Kim Y.S.;	
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays	
RT	and polymorphism."	
RL	J. Clin. Invest. 88:1005-1013(1991).	
CC	-1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND	
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A	
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS	
CC	AGENTS AT MUCOSAL SURFACES.	
CC	-1- SUBUNIT: MULTIMERIC.	
CC	-1- SUBCELLULAR LOCATION: SECRETED.	
CC	-1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,	
CC	Bronchus, Cervix and Gall Bladder.	
CC	-1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR	
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).	
CC	-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND	
CC	VARIES AMONG DIFFERENT ALLELES.	
CC	-1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT	
CC	OF SLIKWORM HEMOCYTIN.	
CC	-1- SIMILARITY: CONTAINS 2 WFC DOMAINS.	
CC	-1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).	







Db 2258 -----STEQPTSTTPGCGQLTFPMNNSSEVLITSEPHVLSSLSLSPDVSGSSTTPNNLSESS 2312

Qy 533 THEPAPTTKEPAPTTPKKAPATTKEPAPTTPKEAPATT----- 572

Db 2313 TVE---TFKTSSEVLSANSEEPSTTEAPPTLSPTDLSTLTNNLSGSSVSTEDRSEISSE 2368

Qy 573 -TRKPAAPAPK-----BPAPTPKEAPAPTPKTLPTTPPELAPTTDEKAPATTP 621

Db 2369 NSKRP-SAPELVTSSVTVAVASSPDVPTES--SEPDDLGSSTENIPFASSKQTISSTP 2425

Qy 622 EELAPTPPEPPTTPPEEP-----APTPKAAPNTPKEAPATT----- 660

Db 2426 TEDTPTASEEPTKSTMSPDLSSTSNVLSSESTTPRESS----KSPVSSSTEGISVITST 2481

Qy 661 --PKPEAP-----TPKEAP-----TPPKAPATTPGTAATTLKEAPATTPKKP 704

Db 2482 ESKSPESTISSVLEEDLTKTTPSPILLETTPASSTSEPTEDSLTVSVRIHELTTSSN 2541

Qy 705 AKPELAPTTTKBPTSTTSDKPA-----PT-----TP--KGAP- 735

Db 2542 VKRESSTTSSSSKPSQEPAGILTSVTVVPTSSVSLTASEIALISNTPFGKGRPI 2601

Qy 736 -TTPK-----EPAPTPKEAPATTPKGTATTLKEAPATTPKKAPKAPKELAPTTKGP 787

Db 2602 TTPSPKSVKSTSPSTVSSSEPSSESTKRTVSTVSTTPPTTEETTSSELITTAAPSKPT 2661

Qy 788 -STTSDKAPATTPKETAPTPKEAPATTPKKAPAPTPPEPTTSEVSTPTTKPTPIH 846

Db 2662 ESTTESSEAPTPPATSTSEKPSNVSTSKSTENETSTSGSGLSESTMSST----- 2715

Qy 847 KRPDESTP--ELSAEPTPALE--NSPKPGVPTTKTPAANK-----PDMTTAKDKTE 897

Db 2716 SEPENAPAVTVYSSSEASTLEENSTSP---TSSEASVLSLFPESITSEAVTNS 2771

Qy 898 RDLATTPETTPAAPMTKETATTEKTESKRTATTT-----QVASTTQ 942

Db 2772 R---APAITMSSSHRISTVSSEPSPELPLSTVSPNVYATSSIPSEPIIISVTS 2827

Qy 943 DTPP-FKITP-----LKTTLAPKVTTTKITTTTETLNNKPEETAKPKDRATNSKATTP 996

Db 2828 SSTPVRILTGPDDLIVSVTVPSHGNNRONTAVSV---PSNSTSPILPSES--LTTPQ 2883

Qy 997 POKPRAKPKPTSKKPKTPMPVRKPK---TTPTRKMTSPMELN-----PTSRILAM 1048

Db 2884 PPTTTTAKKATTTSGKKGPSIOPPAEMFTTPAP-----PPPSNGCYGETINOEEOV 2937

Qy 1049 LQTTT 1053

Db 2938 TSTTT 2942

RESULT 14

151618

nucleolar phosphoprotein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 24-Sep-1999

C:Accession: J151618; S57757

R:Cairns, C.; McStay, B.

J. Cell Sci. 108, 3339-3347, 1995

A:Title: Identification and cDNA cloning of a Xenopus nucleolar phosphoprotein, xNopp180

A:Reference number: J151618; MUID:96019267

A:Accession: J151618

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-990 >CAI>

A:Cross-references: EMBL:X88927; NID:g895920; PIDN:CA61368.1; PID:g895921

C:Genetics:

A:Gene: xNopp180

C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein

C:Keywords: phosphoprotein

Best Local Similarity 24.28; Pred. No. 1.1e-19;  
Matches 246; Conservative 148; Mismatches 424; Indels 199; Gaps 40;

Qy	88	SPSSKAPPPG-----ASQIKSTTRSPRPKKKT-----KVISEBELTEE	13
Db	57	SPDAKKRRPANGLPKKSASKESSSESSSEEDBPAAKQAOPAGCKRPVAAQPKKA	116
Qy	134	HSVSENOESSSSSSSSSSSTIW-IGSKSNAANRELQKKLYKVNKKNT-----KK	186
Db	117	KSSSESSDESEDEETKKPPAKRPAQTPPVAAVPTQKAKSSSSSESSSEDEASKK	176
Qy	187	KPTPKP---VYDEGSGLDNDEKVTVPDSTQHNKVNSTSPKITTAPLINPRLPN	243
Db	177	QVYKVPKQAVYKAGLASNNG---KTDASSSESSSDSPAKKTAATK-----TP-	224
Qy	244	SDTSKETSLTVNKETTVETKETTTTNNKOTSDGKETISAKETOSIEKTSADLAPTS-	301
Db	225	--TKPATAKPAQKATGAKKSSSRSDSSDEDEQTKAKSPKPDVYSAV---PPIVS	279
Qy	302	--KVLAKPPEAKETTKCPALUTTPKEPIPTPKP---PASTTPKEPTPTTIKSAPTTKEP	357
Db	280	KKKTISOPETKAPKPESSSDSSDEEOPAKKAKIVPKAASAAPKLAKAETSTSES	339
Qy	356	ADPT-----TKGAPTPKEPAPTTTKKEPAPTPPKEPAPTTTKE--PAPTTKSA	405
Db	340	DSSSEDEKSSVKLGVAAP--KKAFA-----APDAKSPVAAAKKASAPAKASSSD	300
Qy	406	TPPKPEPAPTPPKKAPTPKKEPAPTPTPKEPIPTPKKEPAPTPKPA--PTTPKEPAPAPK	464
Db	391	SDSSSEETTKPKAKTPAPKASA-ATPISKPTPKGK--ATPISKPAKPGPSTAKDS	448
Qy	465	KPAPTPKEPAPTPPKEPAPTPTKESPPTTKKEPAPTTKSAPTTKKEPAPTTKSAPTT	524
Db	449	SSSDSSSDSEETTKPAAKTTPAKSAAPTSTKPTNSKATPTSKTKPAKPGPKNSA	508
Qy	525	PKESPPTTKKEPAPTPPKEPAPTPPKKAPAPTPKEP--APTPPKPEAPTTKKPAPAPKE	563
Db	509	KKSDSSSDSSDSEKTPA---KRAAKTTPKPAKKTTPAKPA-----AKTPAK	557
Qy	584	PAPTPPKETAPTTPKKLPTTPPEKLAPTTPKPAPTTPEELAPTPPEEPTTPPEEPAPT	643
Db	558	PA-----ASTPKQVPTKKE-----SSSDSSSDSEDEKKS	500
Qy	644	TPKAAPTPKEPAPTPPKEPAPTPPKEPAPTPKETAPTPPKCTAPPTLKKEPAPTPPK	703
Db	551	SAKRPVKT---PGKAVSK--PVVASKPVP-ARKASSSDSSSEETTKTKPLTKLS	643
Qy	704	PAPKELATTTKEPTSTTSD-----KPAPTPKGAPTPPKEPAPTPPKEPAPTPPKG	756
Db	644	PAVKTLPEKKAESSSDSSDSEKTKRPAPKPAKSAPTVNTAKAPAKONKASKASCSDS	703
Qy	757	TAPPTLKEPAPTPPKKPAKPLATTTKGPSTTSDKPAPTTPKETAAPTPKKEPAPTPK	816
Db	704	SSEEGKOP--TGKSPAKATAP-PKPNPAVANKDPSSSSSDSGDEKO---KPK	757
Qy	817	KPAPTPPEPTTSEVSTPTTKEPTTIHKSPESTPELSAEPPTALNSPKEGVPR	876
Db	758	QAA-----AAKVNGAKAAKAPTRPKAASSSE---DS	787
Qy	877	TKTPATTPPEMTTAKOKTTRDRLRTTPPETTTAAAPKATKETATTTKETTESKITATTTQV	936
Db	788	SSDEBVSAAKKTNTNVAKSPV---TPPKAVPAKKKSSSESSDEDEKOGKNTSTTKI	843
Qy	937	TSTTODTTPPKITTLTKTTLPAVTTTK--KITTTIEINKPEETA-KPKDRTATS---	900
Db	844	ANST-----PAAAAECSSESSSEDEGKANGSCKRRKRESTGNAEC	885
Qy	991	KATTPKOPKPAKAKKPTSTIKKPTMPRVKPKTTPTP-RKMTSTMPELNPTSRIAE	1046
Db	886	EAVTPE-----NKKLAKASPTTFRKVNKKEKNTTFRFVVEEDIEIIP--RMD	932

Query Match 8.78; Score 643.5; DB 2; Length 990;

Qy	428	APPTKEPPTTKEPAPTKEPAPT--TPKEPAPTAKKAPAPTKEPAPT--PKEPAP	484
Dy	2153	ALSSPSSSTTESPSPGTTTTCGTRGTSRTTATAPSKRTSTLLPSSPTTAPITTVV	2212
Qy	485	TTTTKESS-----	491
Dy	2213	TTTCEPQACAMSEMDYSYMPPEPSSGDDDTYSNIRAGAGVCEODPLGECRAQAQRCVPL	2272
Qy	492	-----PPTPKKEPAPTTKSAP	507
Dy	2273	RELGOVVECSLDGVLGCMRREQVGKFKMCFNEIRVFCNCYGHCPSPATPSTATPSSTP	2332
Qy	508	TTT--KEPAPTTKSAPTTPKEPSPTTTKEPAPTTPKKEPAPTTPKKAAPTTPKEPAPTTP	565
Dy	2333	GTWWTLLKLTATTTATTTSTGSTATPSSIOGPAGTPIH--VSTTATPTTVYSSK---ATP	2386
Qy	566	KEAPAPTTPKPA-----PTAPKPEA-----PPTPKETAAPTTPKTLTP-	602
Dy	2387	FSSPBGATAPLALRSTATTPTTATSTALPSSSLGTTTWRLSQTTTPMATKMTATAPSSTPE	2446
Qy	603	-----TTPKEAPAPTTPKEPAPT-----PEELAPTTPEEP-	632
Dy	2447	TVHSTVLTATTATTCATGATSVATPSSPTGATTTKVPTTTTGTVMPSSPGTAPRPV	2506
Qy	633	---PTTPEEPAPTTPKAAAPTTPKEPAP--TPKKEPAPT-----TPKEPAPT--TPKE--T	680
Dy	2507	WISTPTPTTSGSTVPSSIPGTHTPVLTLLTTTPQVATGSMATPSSSTQISGTPSLIT	2566
Qy	681	APTTPKKTATPTLKEPAPTTPKPKAPKEAPLT--TPKEPTS-----TTSOKAPAPT	729
Dy	2567	TATTTATGTTT--NPSSTGCTTPPIPELLTATTATTPAATSSVTPSSALGTHHPVPNT	2624
Qy	730	PKGT--APTTP-----KEPAPTTPKEPAPTTPKGTAPT--TLK	763
Dy	2625	TATTHGRSLSPSSPHVKTAMTSATSGTLGTHHTEPSTGSHMPAATGTTTSTPALS	2684
Qy	764	EPAP---TTTKKAPKELATTTKGTSTTSOKAPAPTTPKETAPT--PKBP-----	810
Dy	2685	SPHSSRTTESPPSP---GTTTPGHTTATSKRTTATAPPSKRTISTLLPSOPTSAPTTV	2740
Qy	811	-----	810
Dy	2741	VTTGCEPQACAMSEMDYSYMPPEPSSGDDDTYSNIRAGAGVCEODPLGECRAQAQRCV	2800
Qy	811	-----APTTPKKEPAPTTPETP	826
Dy	2801	LGEIGOVVECSLDGVLGCMRREQVGKFKMCFNEIRVFCNCYGHCPSPATPSTATPSST	2860
Qy	827	PPTT---SEVSTPTTKEPTTHKSDSETPELSNAPPTKALENPKBGPVTK-----	878
Dy	2861	PGTWIILTELTWTTTWTATASGSTATP--SSTP--GTAPPKVLTSPATTPATSSKATSSS	2917
Qy	879	-----TPAPTK-----PEMTTAKDKTTERDLRTTPEPT	907
Dy	2918	SPKRTATLTPVLTSTATSTATSVTPPIPSLIGTGLPQTTTPVATMSTIHPSSTPEPT	2977
Qy	908	---TAAPKMTKETATTEKT-----TESKITATTOVTS--TTTODTP-----	946
Dy	2978	HSTVLTLLTKATTATASSTSPSSPGTWIILTELTAAATTTAGTGPTATPSSSTPCTWIL	3037
Qy	947	KFTITLKTTLAPKVTTT--KTIITTEIMKKPEE---TAKKDRAT--NSKATTPKQO-	998
Dy	3038	TELTATTTATTTSTGSTATLSSPTGPTWILPEPSTTAVTVAPGOSTATVASTTATGTPHV	3097
Qy	999	-----KPTKAKPKPTSTRKKPT---MPRVKPKTTPTPKMT-----	1032
Dy	3098	STTATPTTVTSSKATPSSSPGTAALPALMSTATPTTATSTFALPSSSLGTTWTLSSQT	3157
Qy	1033	-----STMPELNPTSRILAEAMLOTT-----TRENQTPNSKLEVPNKSEDAG-	1074
Dy	3156	TPPATMSTATPSSPERVHSTVLTATTTATGATGAVAPSSPGTAHTTKVPTTTTGGF	3217
Qy	1075	-GABEETPHMLLRPHVMEPTVPMQDYLPRVNOGIIINMLSDENICNGKRPVGLTTL	1133

```

Db      3218  TATPSSSGTALTPPVWVSTTTTPTTTP--TTSGSVTPSSIPGTHIAVLTITTTTV 1184
Oy      1134  RNGTLVAFRGHYFMWLSPPSPARRRTIEWGVLPSLDIVFTKCNCEGKT 1184
Db      3276  ATGSM-----ATPSSS---TQNSGPPSLTTATTTATGT 3309

RESULT 13
T34513
hypotheical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Ravello, A., Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: 221536
A:Accession: T34513
A:Status: preliminary; translated from GR/EMBL/DBJ
A:molecule type: DNA
A:Residues: 1-307 <F&V>
A:Cross-references: EMBL:U013646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Intons: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/1
3504/1

Query Match      8.9%; Score 658.5; DB 2; Length 3507;
Best Local Similarity 22.6%; Pred. NO. 8.7e-20;
Matches 286; Conservative 178; Mismatches 484; Indels 317; Gaps 51.

Oy      3  LSSCAGRCGEGRD-ATC-----NCDYNCQHYECCPDFKRVCTAELSCKGRCFESF 54
Db      1781  MGSCGCKCMAGYTGDCATCIKIEEPRKSDKA----CTDEMSRLCELE--KMCCTVDE 1832
Oy      55  ERGREG-----DCDAOCKKXDKCCPDVESCACAVHNPTSPSSKAP 97
Db      1833  EEVPOGACLPGHHPINGTCQSLQISGLCAQKNDCKNAE--CIDH-----PDSHFCSC 1885
Oy      98  PSG-----ASQIKSTKRSKPPNNKKTKKYLESEIT 131
Db      1886  PDGFIDGMCDVDYECNNAAGMODDENTKCMPTGISFNCVLEGFKKVDEKCVDEKKOP 1945
Oy      132  EEHVSVENQSSSSSSSSSSSTTWIKIKSSKNSANRELQKKLKYKDKNKNTPKKKPPK 191
Db      1946  NRRKIEIDENSSSSNSG-----QEKPTTK 1970
Oy      192  PRVYVDEAGSLGNDGKVTTPDITSTQHNKVKSPKITTAKPINRPSLPPN---SDTSK 248
Db      1971  GIVSSTS-----ATSSSESTTAEPHVHTSISSTSTYKDMSSKSPENVYMSSESP 2019
Oy      249  ETSIVLVNKKETV-ETKETTITNKQISTDGEKKTSAKETQSIIEKTSANGLAP---TSKV 303
Db      2020  EVSTSSSKSTTASSETIVSSTPSSSSSEAPLINSPPATTTVEITESSVSKSTTPKESSSEI 2079
Oy      304  LAKPPTKAEITTKGPAITTPKEPTPTTPKEPASTTPKEPTPTTKS-APTTKREPAPTT 362
Db      2080  TVLKSSKSPVETSSVSKSPSTS-TTQSOVSNIVETSKSVTLSSSEAPVTSISPREVHT 2138
Oy      363  KSAPTTPKEAPPTTTPKEAPATTPTKKEPAPTTTKSAPTTPKEAPPTTPKPKAPAT 422
Db      2139  -SSETKSLSSASTGTGDNSTTPSTISLASVKSTSAPEGTSASAVAVLKSSLSPDVSQPS 2197
Oy      423  TPK-EPAPTTPKPEPTTPKEPAPTTKEP-APTTPEKAPATPAKKAPATTPKEPAPTTK 480
Db      2198  TKTFDATESSTVQASSETISSETSVKSTSEPSHSHTKLKLSINSSSVPTSPSPSTPVE 2257
Oy      481  EPAPTTKESPSTTP--KEPAPTTTKSAPTTTKAPATTTKSAPE-----TTPKEPSPTT 532

```

Query Match 9.0%; Score 666.5; DB 2; Length 7962;  
Best Local Similarity 24.3%; Pred. No. 9.1e-20;  
Matches 268; Conservative 120; Mismatches 491; Indels 229; Gaps 50;

```
QY 74 PDYSFCAVHNHNPSSSSKAPPPSGASQTIKSTYRS-----PKPPNKK 119
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6627 PEEEEFVE--EEVLPRVKRVPVAPVPEIKKKVKVVPKKEAPAKVPEVPEKKV 6884
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 KTKVI--ESEETEEHSVENOESSSSSSSSSSSTIMWIKSSKNSAANRELQKKLVK 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6685 EEKRIILPKKEEVLVPLVYTEPEEPEPISSE-----ETPEEPPITEVEEVAAPRPV 6935
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 DNKKNRKTKKPPKPPVVDAGSGLDN--GDFKVTTP-DTSTQHNKVSPTSKITAKP- 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6936 EVIKKAVPEAPPTPKVPEAKVPKAKSKIIPEEKVPVQVQKKEAPRAVPEPKKVPPEKV 6995
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 -INRPSLP---NSDSKETSILVNKETVEK---ETTTTKQSTQDKE----- 278
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6996 LVPKKEAVPAKGRVLEEVSVAFROEVVYKERLEVEVAEEIPEEEEFHEVEEYF 7055
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 -----KTSAKETQSIKKT-----SAKDLAPTSKVLAKPTPKAETTTKG 317
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7056 EEEGFHEVEEFIKLEQHRVEEHRVENVHVIYFEAEVEVEKPKAP-----KG 7107
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 318 PALT-----TPKEPTPTTP-KEPASTPKPEPTPTTIKSAPTPEK-APTITKSAPPTPK 370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7108 PEISEKIIIPKKPPTKVPKPEKPAKVEPKKIYVEKVRVEEPRVPTKVEVLPRK 7167
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 EPAPTTTKEAPPTPKPEAPTTTKEAPTTTKSAPTTPKAPPTTPK---PAPTTPKE 426
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7168 EVVP-----EKKVPVPAPKAPPEAP-----PKVPEAPKEVEVPKVPVPPKPKPEV 7213
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 PAPTPEPTPTPKPEAPPTTKEAPPTT-----KEPAPTAAPKRAPTTPPEAPATTP 479
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7214 PPTKVEVPKAAVPEKKAAPVAPKIPKESPPPEVEEPEEESPAPK-KPEVPVPRKPEVP 7272
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 480 KEAPPTTKESPPTTKEAPATTTKSAPTTTKEAPTTTKSAPTTPEPSPTTKEAPAT 539
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7273 KEVVPKPKVPAAP--PKKPEVTPVK-VPEAPKEVPEKKVPV--PKKPEVPPTK----- 7223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 540 TPKEPAPTPPKKRAPPTPKPEAPPTTKEAPATTTKKRAPTAPEAPTTTKEAPATTPPK 599
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7324 VPEVPKAAVPEKVP-----EAIPEKESPPPEVEEVEEVALEPEAVEEPEEAPAPQ 7379
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 600 LT-----PTPEKLAPTT---PEKRAPTPPEELAPTTPEEPTP--TPPEEAPPTTPKAAP 650
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7380 VTPPKNPVPEKKAAPVAVAKKPELPVVKVEVPEKVPVLPVK--PEAPKAPV 7437
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 651 NPPEKEAP-----TPPEK-----APTPEKAPPTTKEAPPTTPEKAPPTTKEAPAT 700
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7438 EVPEKEVVPPEKKAAPVPEKPEVPAKVPPEVPPKVP---LEEKPAVP---VPEASESPPEV 7490
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 701 PKKPAPELAPTTTKEPSTTSDKRAPPTPKGAPPTTKEAPATTPKEAPATTPPKGAP 760
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7491 YEE--PEELAP---EEELAPPEEKVPVAEE---EEPEVPPAVPEEKKIIPEKVP- 7540
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 761 TLKEAPPTPKKPAPELAPTTTGTSTSDKRAPPTPEKTA-----PTTPKEAPAT 813
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7541 VPKKPAPEKPEEPEKVP---IEKPKLPRPPPPAPKEDVEKIKFOLKAIPEKKVDE 7597
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 814 TPKEP--APTPEPTPTTSEVS--TPPTTKEP-----TTHKSPDSTPELSAEP 862
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7598 NQVPEKVELTPLKVPGGKKVKKLLPEKKPEKEVVLKSVLRKPEEPEEVE---P 7653
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 863 KALENSPEKGVPTTKTPAATK---EMTTAKDKTTERDQATTPETTTAAPKMTKETA 918
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7654 KMLE-KVKKPAVP---EPPEKPVVEVEVPTTKREKRIPEPTKVEIKPAIDLPAPDEK 7709
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 919 TTTTEKTESKITATTTQVSTTODTTPKITTLLKTTTLLAPKVTTTKKITTTEINMK-- 976
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7710 PKPEA-----EVKTIKRPVEPEPTPIAAVTVAVVGKAAE 7745
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 977 ----PEEKAPKDRATNSKATTPPKOKPTAKPKPTG--TKKPTMPRVKPKTTPTP-- 1028
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7746 AKAPKEAPAKGPIKVPKPTSPISAEKRLKLPGGCKEPPDEAPFTYOLKAVPLKVF 7805
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1029 ---RKMTSTMEELNPTSRIAEAMLOTTT 1053
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 7806 KEIKDILTSESEFVGSSAIFECVLSPST 7833
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 12

T45025  
mucin MUC5B, tracheobronchial [imported] - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T45025  
R:Dessey, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.  
J. Biol. Chem. 272, 3168-3178, 1997  
A:title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter  
A:reference number: 222899; MUID:97166151  
A:Accession: T45025  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3570 <DES>  
A:Cross-references: EMBL:272496; NID:91834502; PIDN:CAA96577.1; PID:91834503  
A:Experimental source: placenta  
C:Genetics:  
A:Gene: MUC5B

Query Match 8.9%; Score 663; DB 2; Length 3570;  
Best Local Similarity 22.1%; Pred. No. 5.8e-20;  
Matches 396; Conservative 119; Mismatches 578; Indels 698; Gaps 72;

```
QY 30 MECCPDFKRYCTLELSCGRCFSEFGRCDCAQCKYDKCCPDYE--SFGAEV-HNP 86
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1581 VECSLDFGLVC-----RNRE-----QVGKF-KKCFNVEIRFCNCVGHCP 1619
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 TSPSSSKKAPPS-----GASQTIKSTYRSKSPKPKKKTKKVI 126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1620 STATSTATPSSPTPGTWTLLTBOITAAITTAATGTAIPSSPTGTAAPP-----KVLV 1673
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 SEETIEHSVENOESS-----SSSSSSSSSTIMWIKSS---KNSAANR--- 168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1674 SQATTPATASSKATSSSSPTATTLPLVLTSTAKSTATSTPTLPSSTLGTGTGSOHRPH 1733
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 -----ELOKILVNDKNNKRIKKKPEKPP-----VUDE-----AGS 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1734 PMATMSTIHPSSTPTHTSTVLTAKATTTRATSSMSSTSPGTTWILTETTAATTA 1793
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 201 GLDNG-----DEKVTTPDTSTYOHN-----KVSSTPKITTAKP 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1794 ALPHGPRSSPGTWTLLTEBSTATVTVPRGATATASSRATAGTLKVLTS-TATTPPYI 1852
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 NRP-----PSLPNSDSKETSILVNKETVEK---KETTTNKQSTDG 276
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1853 SSRATPSSPGTATLPAKSTATPTATSVTAIPSSSGTATRLSQTTTTPATMSTAT 1912
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 277 KEKTSAKETQSIKETSADIAPTSKVLAKPKPKAKATTK-----GPA 319
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1913 PSSPTPTVHTSTVLTATTTTRIGSVATPSSSPGTATHTTKVPTTTTGTATPSSSPGTA 1972
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 LTPP-----KEPPT-----TPKE-----PASTTPKEP 344
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1973 LTPPVWISTTTPTPTRGSTVTPSSIPGTHATAVLTITTTTVAATGSMATPSSSTQTSR 2032
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 TIKKSAPTPPKKEAPTTTTSAP--TPPKPAPTTP-KEAPAT---TPKEAPATTTKEPA 397
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2033 PSLTTATTTTATGTTTNPSSPTGTPPIPPVLTATTAATAASSVTTPSSALAGTHTTPV 2092
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 398 PTTTSA-----PTTP-----KEAPPTPKKPAPTT---PKPE 427
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2093 PMTATATHTHSLRPSSPHIVPTAMWTSATSGILGTHITPEPSTGTSTHTPATTTGTQPS 2152
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Biochem. Biophys. Res. Commun. 165, 644-649, 1989

A>Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated A:Reference number: A36735; MUID:90088473

A:Molecule type: mRNA

A:Residues: 1-142,'O','144-162,'Q','164-168 <ABE>

A:Cross-references: EMBL:M31823; NID:g181543; PID:AAA35757.1; PIG:g181543 R:Kusuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H.

J. Biochem. 112, 609-615, 1992

A>Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglutinin binding site in human breast cancer membrane proteins

A:Reference number: JX0235; MUID:93123189

A:Accession: PX0066

A:Molecule type: mRNA

A:Residues: 998-1011,'ES','1014-1017,1018-1032,'T','1034-1037,1038-1057 <MAS>

A:Experimental source: gastric carcinoma cell

R:Zilhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Kiydar, I.; Wreschner, D.H.

FEBS Lett. 356, 130-136, 1994

A>Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine n A:Reference number: S51026; MUID:95080414

A:Contents: annotation

A>Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region are C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 C:partial repeats. The repeat shown is defined by SmaI nuclease sites.

C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146

C:Genetics:

A:Gene: GDB:MUC1; PUB

A:Cross-references: GDB:120705; OMIM:158340

A:Map position: 1q21-1q23

A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3

C:Superfamily: polymorphic epithelial mucin

C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polysphism F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>

F:1-62/Region: mucin 1 amino-terminal non-repetitive

F:1-23/Region: signal sequence #link PREA #status predicted <SIGA>

F:1-19,-29-32/Domian: signal sequence #link PREB #status predicted <SIB>

F:1-19,-29-324/Product: mucin 1 precursor, splice form B #status predicted <PREB>

F:1-19,-29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice F:118-1017/Region: 20-residue repeats (GSPAPAHGVTSAPDRPA)

F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetive

F:1245-1272/Domian: transmembrane #status predicted <TRM>

F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status prediced F:1123/Binding site: phosphate (Tyr) (covalent) #status predicted

Oy 518 TKSAPPTPKESPPTTMEP-----APTTKEPAPTPK----KPAPTTKEPAPTPK- 566  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 368 T-SAPDT--RPAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRA 424  
 Oy 567 ---EPAPPTTKKPAATPAKEPAPTTKEATAPTTKCLPTTPEKLABTTEKPAPTPE- 622  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 425 HGVTASAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRA 484  
 Oy 623 ---ELAPTTPEEPPTTP-----EEAP--TTPKA----AAANTKEPAPTPK- 662  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 485 HGVTASAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRA 544  
 Oy 663 ---EPAPTTKEPAPTTKEATPTTPKGTATPTTKEAPTT-----KKAPAKEL 709  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 545 HGVTASAPDTPRAPGSTAP---PAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPGST 600  
 Oy 710 APPTTKEPTSTSDKRAP--TTPKGTAPTTKEPAPTTKEPAPTTKEG---TAPTTLK 763  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 601 AP-PAHGVTSAPDTPRAPGSTAPRAHGVTS---APDTPRAPGSTAPRAHGVTSAPDTPR 655  
 Oy 764 EPAPTP-----KKAPAKELAPTTTGGPTSTSDKRAP--TTPK----ETAPTP 807  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 656 APGSTAPRAHGVTSAPDTPRAPGSTAP--PAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTR 714  
 Oy 808 KEPAPTTP-----KKPAPTTPEPTPEPTTSEVSTPTTKEPTT-----IHKSPD 850  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 715 PAPGSTAPRAHGVTSAPDTPRAPGS--TAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPD 772  
 Oy 851 ESTPELSAEPPTKALENSPKEBVPPTKTPA-----TKPEKTTAKK---TTERD 899  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 773 TRPAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPD 832  
 Oy 900 LRTPEPT-----TAPKMKKETATTTKTESKINATTOYSTTTOPTT--FKIT 950  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 833 TRPAPGSTAPRAHGVTSAP-----DTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGV 888  
 Oy 951 TLKTTTLLAPKVTTTTKTTTT--EIMNKPETAKPKDRATNSKATTPKQKPTAKPKPT 1008  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 889 SAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAP--GSTAPRAHG 946  
 Oy 1009 SRRKKKTPRVRKKPTTTTPRKMTSTPELND---TSRIEAMQI--TTRP---NOTPN 1060  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 947 VTSAPDTPR--APGSTAPRAHGVTSAPDTPRAPGSTAPRAHGVTSAPDTPRAPGSTAPR 1003  
 Oy 1061 SKLIVVNRKSEDPAGAEETPRMLLRPHVFMPEVTPDMDYLPKRVNQGIIINPMLSDETN 1120  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 1004 AHGVTSAPDTPRAPGSTAPRAH-----GVTSAPDNRPALGSTA---PVAHNYS 1049  
 Oy 1121 IONKRPVDGLTLRNGCLVAFNGHIFWMLSPSPS 1156  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 Db 1050 ASGSASGSASTLVHNGTSARATTTTPASKSTPSPIS 1085  
 ||| : : : : ||| : : : : ||| : : : : ||| : : : :  
 RESULT 11  
 158346  
 elastic titin - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
 C:Accession: 158346  
 R:Label: S. Kolmerer, B. Science 270, 293-296, 1995  
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity  
 A:Accession: number: A57430; MUID:96026530  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-7962 <RES>  
 A:Cross-References: EMBL:X90569; NID:g1017426; PID:CA662189.1; PID:g1017427  
 C:Genetics:  
 A:Gene: GDB:TTN  
 A:Cross-References: GDB:I27867; OMIM:188840  
 A:Map position: 2q31-2q31

```

OY 49 RCFSEFEGRECDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKAPPPGASGOTIKST 108
Db 34 RDKFVKKHKKMTCTSC-----KCPD-----APSN-----PBDVSTIIS- 70
OY 109 TKRSPKPNKKTKKVISEETIEHSVSENOESSSSSSSSSSSTIKKISSKSNANR 168
Db 71 -----LNDVNDIGPSGDSNPTGSSMFOEIAEVVGGOTVSEHNIDSSV 114
OY 169 ELQKKLVKKNKKMTKPPKPPVNDGASGLDNGDFKVTTPDSTTQ---HNKSTS 225
Db 115 EVEKKV-----TTSIDASTTNAPTGKSDST 140
OY 226 PKHTA-KPMPRSLPNSDT-----SKETSLVKNKETTVEKETTNNKQSTDGK 277
Db 141 PEITGIVVINSKESVYDMSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 199
OY 278 EKTTSAKETOSIEKTSADKLAPTSKVLAKPPKAEITTKGALTPKEPPTTKPEAST 337
Db 200 TEIASPMETINTTEATTSVSPTSILAS---EDETVTAIAEST-----TVIAEVST 250
OY 338 TPKEPTPTTISAPTTKREPAPTTTKSAPTTKREPAPTTTKREPAPTTTKKEPA 397
Db 251 TTEPPTT-----AESITTKSTT---KAPATTEEPPTTEE---VTTTBAET 292
OY 398 PTTKSAPTTKREPAPTTT---KKPAPTTKREPAPTTKREP---TPPTKREPAPTTKREP 452
Db 293 STTSSSESTSEK---PTTPLIDNKIAPATGK---PETTHFPVGTTPN---FDTAETTP 344
OY 453 TTPKEPAPTLAKKAPPTTKREPAPTTKREPAPTTTKE-----PSPTTKREPAPTTKSA 507
Db 345 VANSSEDMKTLSTKMAETTQOTTEVT-DGPEKETTKVNSIEITTPVLETTSTSTAS 403
OY 508 -----TTTKREPAPT---TTSAPTTKREPAPTITKE---PAPT-----TPK--- 542
Db 404 KESDGEFTTILAKLVYTAADSDSTESATVWKPNEETTKSHVWPKKGVKVPKLELS 463
OY 543 --EPAPTTKPPAP-----TPKE-----TPKE-----PAPTT 565
Db 464 FDEPTTEIT-KAPHPGKLEKTYHFLVLSDFANFVSEKENDDNHLDYNNREAKERTTT 522
OY 566 KEPAP-----TTTKKRAPTA---KREPAPT---KREPAPTTKPLPTPEKLA---PTTP 613
Db 523 EESSTTEEVVTEEPANGTNPPTTEPTTEPTSTAESTTTALPPTTEQVTEBPTTA 582
OY 614 EKRPAPTTPELATTEEPPTTREPAPTTTTPKAAPNPKREPAPTTKREPAPTTKEPA 673
Db 583 EKSTANO---KPTTQESYST---EKSTTKKA---STTEE---PTTDEPTTPTT---BS 627
OY 674 PTPKETAAPTTKGIAPTTTLKEBPAPTTKPPAKELAPTTTKEPSTSTSDKRAP-----TT 729
Db 628 STTGKATPELSTSTSEETTTTELKITE-----GSTTEBPTTTALFAASIGITTT 679
OY 730 PKGTAPTTKREPAPTTKREPAPTTKGIAPTTTLKEBPAPTT----- 770
Db 680 DEETSTTSTTEPITSTK---IVTESALITQTSVSVESSTPQOLBERMKVAINKEKHNL 737
OY 771 ----KKRPAKELAPTTTKGP-----TSTTSKRAPPTPKETAPTTKREP 810
Db 738 VLKEKKRLLKRESSTGSDSEPTTVAENIDEVTTTEKKKVVQTPPTTEKSTOEE 797
OY 811 APPTKPPAPTTPE---TPPTTSE---VSPTPTTKETPTTHKSPDESTPELSAPPTKALEN 867
Db 798 TTTTTEKTTSTKTTTEKPTTSASATETTTSEPTST---TEST--- 838
OY 868 SPKEGVPPTTKPAATPKPMJTADKIDTERLARTTPEPTTA----- 909
Db 839 -----TVDISATTEESSSTAETTTTSAE---TSETTSSASAFTTGSPENTALQS 887
OY 910 -----APKMTKEPATTTTEKTTESKITAATTO---VTST---TODT 944
Db 888 SSOQSEENESSAEKPGARDFVPRKHKHTTVKPAETTSVAASSTTTTEITTEKSTTLET 947
OY 945 TPFKITTLTKTLTAPKV---TTKTTITTTTEIMKPEETAPAKRDRANVSKATTPKPO--- 998

```

```

Db 948 TPTEATTLNEVTPGAPETVGPVDETTINTLELISK-----INTTOISQPKPTDIS 997
OY 999 -----KPYAKAKKPTSTKPKPTMPVRKPKPTTPPKMTSMPLN----- 1039
Db 998 KTDALSSLSLGLSGFTKAPMARTI-----HTTDAFVTAETASLNDGSDK 1044
OY 1040 -----PSTRIAEA 1047
Db 1045 KIIDEAQPTDEIRRA 1059

RESULT 10
A35175
Mucin 1 precursor, repetitive splice form A [validated] - human
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein K939; epista
ncretic mucin; polymorphic epithelial mucin (PEM)
N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Jun-2000
R:Rigdenberg, M.J.L.; Vos, H.L.; Gemmessen, A.M.C.; Hilkens, J.
J. Biol. Chem. 265, 5573-5578, 1990
A:Title: Epistatin, a carcinoma-associated mucin, is generated by a polymorphic gene
A:Reference number: A35175; MUID:90202794
A:Accession: A35175
A:Molecule type: mRNA
A:Residues: 1-952,1033-1344 <LIG1>
A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124;
A:Experimental source: splice form A
A:Note: Genbank entries HUMEPIS1 and HUMEPIS12 present only the amino and carboxyl
A:Accession: B35175
A:Molecule type: mRNA
A:Residues: 1-19,29-952,1033-1344 <LIG2>
A:Cross-references: GB:M32738; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129;
A:Experimental source: splice form B
A:Note: Genbank entries HUMEPIS1 and HUMEPIS12 present only the amino and carboxyl
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Butc
J. Biol. Chem. 265, 15286-15293, 1990
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epith
A:Reference number: A35886; MUID:90368715
A:Accession: A35886
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-992,1033-1344 <GEN>
A:Cross-references: GB:J05581; NID:g188869; PIDN:AA59876.1; PID:g188870
A:Note: Genbank entry HUMMUCAB includes one copy of the tandemly repeated sequence
R:Alan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716
A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-1109, 'S', 1111-1339, 'A', 1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
A:Note: Genbank entry HUMPANMU contains four fewer copies of the tandemly repeated se
Eur. J. Biochem. 189, 463-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen
A:Reference number: S10571; MUID:90276413
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19,29-1155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1344 <HRE>
A:Cross-references: EMBL:X52229; NID:g37053
R:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19,29-1155, 'P', 157-175, 'P', 177-182, 'A', 184-212, 1033-1037, 'A', 1039-1344 <
A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
R:Abe, M.; Siddiqui, J.; Kufe, D.

```



Query Match 10.9%; Score 809; DB 1; Length 1367;  
Best Local Similarity 27.5%; Pred. No. 2,3e-26;  
Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53;

QY 61 DCDACCKYK---CCPEYSECAEVNHPSPSSKAPPPSGASQTIKTKRSPKPN 117  
DB 174 DLTGCTNNYDNOGHSQTFEPGFYWNIDCDNNCGCTSTSTSSSES---STT----- 223  
QY 118 KKTTKVIESEITEEHSEVENOSSSSSSSSSSSTTWKTKSSKNSANRELKLVK 177  
DB 224 ---TSSTSESTTSSSTSESTTSSSTSSSTT-----APAPPTTSCREKPTP 274  
QY 178 DNKNRTKAKTPPKPVVDDEAGSLDNGDFKVTTPDTSTTQHNKVTSPKITTAKPINR 237  
DB 275 PTTTSCREKPT--PPHND-----TTPCT-----KKTITSTCTCKITTTVP 314  
QY 238 PSLPPNSDTSKETSLYVKNKETTETKETTNNKOTSDGKEKITSAKTOSIEKTSKDL 297  
DB 315 PT--PSSSTTSSS-----APVPTPSSSTT-----ESSSAPVTSSTSSSS--- 353  
QY 298 APTSKVLAKTPKAETTTGPAITTPKEPTTPPKEPASTTPKEPTTITKSAPTPKPEP 357  
DB 354 APV-----PTPSSSTESSAPVTSSTESSAPVTSSTESSAPVTPPSSSTESS 407  
QY 358 APTT-----TKSAP-----TPKKEPAPTTTKEPAPTTKEPAP-----TTKEPAPTTTSA 404  
DB 408 APVTSSTESSAPVTSSTESSAPVTSSTESSAPVTSSTESSAPVTPPSS 463  
QY 405 PTPKKEPAPTTPKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 461  
DB 464 STESSASAPVT---SSTESSAPV---PTPSSSTESSAPV---SSTESSAPV 512  
QY 462 APKPAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 521  
DB 513 -----PTPSSSTESSAPVTPS---SSTESSAPVTPS---TSSSSAPVTPPSS 559  
QY 522 PTPKKEPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 580  
DB 560 STESSSTPTSS---TSSSSAPVPTPSSSTESSAPVPT---PSSSTESSAPV 611  
QY 581 PKKPAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 639  
DB 612 PTPSSSTESSAPVTPS---SSTESSAPVTPS---PTPSSSTESSAPVTPPSSSTESSA 666  
QY 640 PAV-----TTPKAAAPNTPKKPAATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 693  
DB 667 PVPPTSSSTESSAPVT---SSTESSAPVT---SSTESSAPVTPPSSSTESS 718  
QY 694 KEKAPPTPKKPAKELAPTTTKEPAPTTTSDKPAATTPKGAATTPKEPAP---TPKPA 750  
DB 719 SAVPPTPSSSTESSAPVPTPS---SSTESSAPVTSSTESSAPVPTPSSSTESSA 777  
QY 751 PT--TPGTAATTLKPAATTTPKKPAKELAPTTTGGPITSTSDKAPATTPKETAPTTKE 809  
DB 778 PVPPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSSNTTSSAPSSPTPSSSTESS 837  
QY 810 PAV-----TTPKPAATTPETPTTSEVSTPT-----TTPKEPTIHKSDESGTPELAE 859  
DB 838 PVPPTPSSSTESSAPVSSSTESSAPVPTPSSSSNTTSSAPSSPT---PSSSTESSST 894  
QY 860 PTPKALENPKKEGVPPTTKTPAATPEKMTAKKDKTTERDLKTPETTTAAPKMKETAT 919  
DB 895 GT--TVPPTSSSKTPGSGOTELSVSTETETIVPKTTTSTVTSTTTTTCSTGNSAGE 953  
QY 920 TTEKTESKITAT--TTOVSTTTQDTPPKITTLKTTLPAKVYT---TKKTTT---EI 973  
DB 954 TTSKGSPTVTTVPTTTTSTTTTSTTTTTCSTGNSAGETTSKGSKTTTITTVPC 1013  
QY 974 MKKPEETAKPKDRATNSAATTPKPKP-----TKAPKKTSTKKPKETM 1016  
DB 1014 STSPSETA-----SESTTSTPTPTVTTVSTTVTTEYSTSKPGGELTTTFFVKN 1065

QY 1017 PVRKPKPTTTPPKKMTSTMBELNPTSRIAEAMLOTTTRPNOTPNSKLVEVNPKSEDA 1076  
DB 1066 PTTVLTITTAAPT--SVTATVNFPTT--ITTVGSCST-----NSAGETTSKC 1109  
QY 1077 EGEPHMLLRPHFMFMEVPPDMOYLPRVNOGIIINPMUSDENTICNGRPVDGLTTLRNG 1136  
DB 1110 SPKTVTTTTPCSTGTEGYTETATLTAVTTVVTTESTGTNSA--GKTTTGYYTKSV 1168  
QY 1137 TLVAFRGHYFWMLSPPSPSPARITREWGIPSDIDVPTRC-----NCCKGT 1184  
DB 1169 TT-----YVTLAPASAVTPATN-----AVPTTITT--TECSAATMAGET 1207

RESULT 7  
S49915  
extension-like protein - maize  
C:Species: Zea mays (maize)  
C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
R:Accession: S49915  
R:Rubinstein, A.L., Broadwater, A.H., Lowrey, K., Bedinger, P.A.  
submitted to the EMBL Data Library, June 1994  
A:Description: Pex genes: pollen-specific genes with extensin-like domains.  
A:Reference number: S49915  
A:Accession: S49915  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1188 <RUB>  
A:Cross-references: EMBL:Z34465; NID:9600117; PIDN:CAA84230.1; PID:9600118

Query Match 10.6%; Score 784; DB 2; Length 1188;  
Best Local Similarity 29.9%; Pred. No. 2,1e-25;  
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

QY 168 PTPKPVVDDEAGSLDNGDFKVTTPDTSTTQHNKVSQPKITTKPINPRESL-----PPN 243  
DB 460 PTPHSPPAD-----DYVPTTPPVGKSPATSPSPQVQPPAASPPSLVLKLSBPQ 510  
QY 244 SDTSKETSLVKNKETTETKETTNNKOTSDGKEKITSAKETOSIEKTSKDLAPTSKY 303  
DB 511 APVG-----SPPPVKTTSPAPIG-----SPSPPPVSVV 541  
QY 304 -----LAKPTKAEITTKGPAITTPKEPTTPPKEPASTTPKEPTTITKSAPTPKPE 358  
DB 542 SPSPPVKSPPPAPVAGSPPEKSPPPAPVAVSPPPVKSP--PPITLAVSPPPVKSP 599  
QY 359 PTTTKSAPTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 415  
DB 600 PPAVAVSPPPVKSPPTPVASPPPAVASSPPPKSPPPPTPVSSPPPEKSPPP 659  
QY 416 PKKPAATTPKEPAPTT-----PKE-----PTTPPKKEPAPTTKEPAPTTKEPAPTPAKKP 466  
DB 660 PAKSTPPEEVPPTPTSVKSSPPEKSLPPTLIPSPPOEKPPTPSTPKP--PSSPEK 718  
QY 467 APITPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 526  
DB 719 SP--PKKEVSSPQTP---KSSPPAPVSSPPTPVSSPPALAPVSSPVKSSP--- 768  
QY 527 EPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 586  
DB 769 PPAVLSPPPAQVKKSPPVQVSSP--PPAPKSSPLAPVSSPPQVEKTSPPPAVSSPP 827  
QY 587 TTKETAP-----TTP-----KKLTPTPEKLAATTPKEKPAV-----TTPBELAATTPDEEPT 633  
DB 828 LAPSSPPAHVAVSSPPVAVSSPPPAVSSPPLTPKASPAHVAVSSPEVAVKSP--PA 885  
QY 634 PTPPEEAPTTPKAAAPNPPEK--PAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 692  
DB 886 PTVVISP--PSEPKSSPPPTVSLPPTIVKSSPPPAVSSP--PMTPKSSPPPVAVSSPPT 943  
QY 693 LKE---PA-----PTTPPK---PAKELAPTTTKE---PTSTTSDKPAATTPKGAATTP 738  
DB 944 VKSSPPPAVAVSSPPATPKSSPPPAVNLPPPEKSSPPPTPVSSPPPA--PKSSPPAP 1000





```

Db 988 ETTAAPEETMYAPTEETTYAPAEETPYEPTTEETTYAPAEETTYAPTEETTY 1047
Oy 761 -----TLKEPAPTPPKP-----APKE---LAPT---TKGPTSTSDKAP 796
Db 1048 ASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTE 1107
Oy 797 TTPKETAPTPKEBAPTPPKAPPTPEPTTSEVSTPTTKKEPT--TIKSGPDESTP 854
Db 1108 ETTAAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPTEETTYGPTTEETTY 1166
Oy 855 ELSAEPKPA-LENSPKERG-----VPTTKPAATKDEMTTAKDKTERDLRTTP--- 904
Db 1167 YAPTEETPYAPTEETTYEPGEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1226
Oy 905 ETTAAAPKMTKETATTEKTESKITATTQVSTTODTT-----PEKITP--- 951
Db 1227 EETTYAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETMYAPIDE 1281
Oy 952 LKTTTLAPKVTYTTKTTTTEIMNKPPEETAKPKDRAT--NSKATTPKP-OKPTAKPKPT 1008
Db 1282 TEETTYAPTEATYAPTEETPYAPTEETTYEPTEETTYAPTEETTYAPTEETTYAPMEET 1341
Oy 1009 -----STKRP-KTMPRVKRPKTTPTPKMTSTWPELNPTRISAMQTTTRP 1055
Db 1342 PYEPAESTVSTKPCNTBEPTDEPTDEPTDE--PSEPTDEPTDEPTDLPTDEPSTP 1399
Oy 1056 --NOTPNSKLVENPKSEDAG 1074
Db 1400 CDNOGINGIGVENKVRYNMAG 1420

RESULT 4
T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leinbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 11274 <LEI>
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GNOC
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CSP:F35A5.1
A:Map position: X
A:Introns: 1272/2

```

```

Query Match 11.4%; Score 846; DB 2; Length 1274;
Best local similarity 27.6%; Pred No. 6.5e-28;
Matches 313; Conservative 119; Mismatches 455; Indels 246; Gaps 60;

```

```

Oy 90 PSKKAAPPSGASOTIKSTTKRSP---KRP---NKKTKKVLSESEITEHSVSENQES 142
Db 185 PSFKKAAPSEKNDPIVPPIPIKPKAKKMPWEDDEVPEETKEPAPTRAKPKALKKRP 244
Oy 143 SSSSSSSSSSTIMIKSSKNSAANRELOKILKVKDNK-KNKKKK-KP-----TP----- 190
Db 245 STSVKAVSDPSPTKKV-----PVKKEPEVPPTPIKNPKKKRPWEDTEPVEEVK 294
Oy 191 KPPVNDAGSGLDNGD----FKVTPDSTQTNKVKSTSPKITTAKPI--NPPRSLPPNSD 245
Db 295 EPPVPEKKAAPVLKKKPDAPAAKADSPSKAPKAPKVEPSPVPPPVANPKKKKPPME 354
Oy 246 TSKETSLLYANKETVETKETTNNKOTSDGKEKTTSAKETOSIEKTSKADLAPTSKYLA 305
Db 355 VDDEPPEEVKKKSAPKPKPVLLKRRKEPPSSSTPPSSDPSPKAAAVKPRDSDPKKATPL 414

```

```

Oy 306 KPIPKAE-----TTTKGPA-----LTPKEPPTTPP---KEPASTTPK 340
Db 415 QADPKAQEVPPTPKVKNVKKRPPWEDDEDVPEEVQPEAPAKKTPVLKRRKEPAKDN 474
Oy 341 EP-----TP--TTSAPPTPKAPATTTKSAPTPPKAPATTTKKEPA----- 381
Db 475 KPAATSKPTPEPKKDPVKPDSSEKKVAAKPSDAQAA--TPVKNPVKKRPPWEDDETPA 533
Oy 382 -----PTPKREPAPTTKEBAPT-----TKSAPTTKEP-----APTTPKAPATT 424
Db 534 DVSKPPTDAKKTSLAKDAPAKESLKPADKADAKAPRDEPKKVAPTAPAKKTPVLA 593
Oy 425 KE-----PAPTPKEPPTTPPKAPATTKEBAPTTPK--EPAPTPAKKAP-----TPKE 473
Db 594 KKEBAGPADSKTEPEKSPKRDSPKAVPAKVPPTTEVAAPVKKKEPTISKRDPAKPK 653
Oy 474 PAPTPKEBAPTTPKEB-----SPTPKREPAPTTTKSAPTTTKREPAPTTTKSAPTPP-KEP 528
Db 654 AEPNSPVVP-PTVKNPVKKRPPWEDDDAPAKPVSLPEEKK--TVVLAKKAPTKDSEA 711
Oy 529 SPTTKEBAPTTPK--EPAPTTKKAPAP--TPPKREPAPTTPKREPAPTTKKAPPA-PKE 583
Db 712 AADPVSGPSSKDKLAKKAPVKRDEPSPMKAAVPIKPAKT--EVPPAVVKKKEPVAKSD 769
Oy 584 PAPTPKEBAPTTPKLTPTTP-----EKLAPTPKEBAPTTPPEELAPTTPEEP-----T 633
Db 770 PSKKAAP-AEPNSP--VPEPTPVKNPVKKRPPWEDDDAPAEVNVPEEKKTPVLA 826
Oy 634 PTPKEBAPTTPKAAAPNTPKEBAPTTPKEBAPTTP--PKREPAPTTKETAPTPPKGAPTT 692
Db 827 PVKPRDPSPKAAVPAKPSIKTDAPVSVKKPEVSKRKEPSKKAEPNSPVVP-----PTP 882
Oy 693 LKEBAPTTPKPKAPKELAPT--TKKEPTSTSDKRAPTTPKGAPTTPKREPAPTTKEBAPT 751
Db 883 VKNPVKKW-KPWEDEDETEEVKKPSE--PEKKTPIVLAK-KEPEKPD-APKVAAPRD 937
Oy 752 TTPKGAPTTLKEBAPT-----TPKKRAP-----KEIAPTTTGGPSS-----TTSK 793
Db 938 PSFKKAVPE--KEPAKVAAPRDLSPKKAIPANTQGEAPTPVKNPVKKRPPWEDDE 995
Oy 794 P-----APTTPKET-----APTKEBAPTTPPKRAPTPPTPEPTTSEVSTPTTKRP 842
Db 996 PAEPVSAPEPEKKTPIVLAKKAPKPRD--SPKKAAPVAAK-PDKPIEV-PPIPVKNP 1050
Oy 843 TTIHKSPP-----DESPLESA-EP---TPKALNSKEPEV-----PTTKTPA 881
Db 1051 VKKMRPPWEDDEPSPVSAPEPEKKTPIVLAKKAPTKPATKPSDAADPVSGPTSKDPK 1110
Oy 882 AT-----KPEMTTAKDKTERDLRTTPPETT--TAAPKMTKETATTEKTESKITATT 933
Db 1111 LSKKAPVEKRPRTTDPKDKLKPSPAKKPEKAPDEPAARKKKAPVWDDDEDEADFTVPA 1170
Oy 934 TOVSTTODTTTPPKITTLKTTTLAPKVTYTTKTTTTEIMNKPPEETAKPKDRATNSKAT 993
Db 1171 PSKKPDTEDEPADPLG-----GPKTDPK-----LKKKAPAEKPREK----- 1206
Oy 994 TPKPOKPTAPKPKPTSTKKPKTMPRVKKPK-----TTTPPKKMTSTMP 1036
Db 1207 -PKPEVSKPEPKPIEPPKP-AAAPKKMRPPWEDDDEPDEADFTMPAPKPKPTEDP 1259

```

```

RESULT 5
t30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yotsov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a mus
A:Reference number: Z20889; MUID:96312450

```



J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.  
A:Reference number: A61257; MUID:91086481  
A:Accession: A61257  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 191,192-1948,1952-1954 <JAN>  
A:Experimental source: bronchus  
R.Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, A.; Blochem. Biophys. Res. Commun. 183, 821-828, 1992  
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus  
A:Reference number: P00328; MUID:92198477  
A:Accession: P00328  
A:Molecule type: mRNA  
A:Residues: 2328-2468 <XUG>  
A:Cross-references: GB:M6523  
A:Experimental source: small intestine  
A:Accession: P00329  
A:Molecule type: Protein  
A:Residues: 2328-2342, K, 2344-2354 <XUG1>  
C:Gene: GDB:MUC2  
A:Cross-references: GDB:120203; OMIM:158370  
A:Map position: 11p15.5-11p15.5  
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von F;2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>

Query Match 14.78; Score 1090; DB 2; Length 3020;  
Best Local Similarity 27.48; Pred. No. 1.4e-37;  
Matches 378; Conservative 107; Mismatches 514; Indels 382; Gaps 51;

QY 19 TCNCDY--NCOHYECCDPFKRVCATLSCGRCFESFERGECDDACQKRYDKCCPDY 76  
Db 1079 SCSDDTGDECECFCSAVASAEQECTKEGAC-----VFMTPLD-CPIFCDDYN---PRH 1128  
QY 77 ESFCAEYHNPTSPSSKAPPPSGASQTI-----KSTTKSPK--PNKKTKKVESEE 129  
Db 1129 E--CEMHYECGNSEFETCTINGIHNSISVLEGCYPCPKDRPIYEDLKCCTADK 1186  
QY 130 -----ITEEH-----SVSNOSSSSSSSSSSSTIWKIKSSKNSAANRELQKIKLVADNKK 181  
Db 1187 CGCYVEDTHPPGASVYETFECKSCVCTNSQVVCPRBECK-----IL 1229  
QY 182 NRTKKRPPTPKPVVDAGSLDNGDF--KVTPPDISTQH--NKVSTSPKTTAKPIN 235  
Db 1230 NQOQ-----DGAFCYWEICGPNGTVEKHFNICSIITRPSLTITFTTI 1271  
QY 236 PRSLPNSDTSKETSITVKKETVEKET--TTNK-----QTSDDGKE 278  
Db 1272 TLPTTP-----TSFTTTTTTTTTPTSSVLTSTPKLCLCWSMDWINEHPSGSDGDR 1323  
QY 279 KITS--AKETOSIEKTSARD----- 296  
Db 1324 EPPDVGCAPEDEICRSVADPHLSLBOHOKVOCDSVGFICKNEDQFNGPFGLCYDK 1383  
QY 297 -----LAPTSKYLAKPTPKAETTTGPAITTKKEPTTPPKAPASTTKKEPTP 344  
Db 1384 IRVNCMPMDKCTITPSPTTPSPPTTTTTLPTTSSPTTTTTPPTTTPPTTTPSPPT 1443  
QY 345 TTKKSP--TTKEPAPTTTKKSAPTPKKEPAPTTTKKEPAPTTTKKAPATTKS 403  
Db 1444 TTTTTPPTTTPSPPTISTTTTTPPTTTPSP--PTTSSPTTTTSSPT--TTTTTTPPTTTPS 1501  
QY 404 AP--TTKEPAPTTPKKAPATTKKEPAPTTPKKEPTPTTKKEPAPTTKEPAPTA 462  
Db 1502 PPTTITPTTTPASTTTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTP--PTTTLPTPT 1560  
QY 463 PKKAPTTTPKEPAPTTPKKEPAPTTTKESPTTPKKEPAPTTTSAP--TTTKKEPAPTTTSA 521  
Db 1561 TSPSPPTTTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPSP 1620

QY 522 P--TTKEPSPTTTKKEPAPTTPKKEPAPTTPKKAPATTKKEPAPTTTKKAPTA 580  
Db 1621 PTTTPTTTPPTTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTTTPPTT 1680  
QY 581 PKKAPTTTPKEPAPTTTKKLPPTTPEKLPATTEPEKAPATTEPEKAPTEPEPTPEEP 640  
Db 1681 PSSPTTTP--SPPTT--TMTTTPS--PTTSSPTTTPPTTTPSSPTTTP--SPPTTMTTP 1731  
QY 641 APPTPKAABNTPKKEPAPTTPKKEPAPTTPKKEPAPTT--KETAPTP-- 685  
Db 1732 SPPTTTPPTTTPPTTTLPTTTPSSPLATTPLPSPITPTPSFSTPTTPPTPCVLCNMWGT 1791  
QY 686 --KG----- 687  
Db 1792 LDSGKPNFHKRGDTLIGDVCSPGMANISCRATWPDVPIGOLQGYVCDVSVGLCK 1851  
QY 688 -----TAPTLKAPATTPPKKAPKELAPTT--TKRP 717  
Db 1852 NEDQKRGCVPMAFCLNVEINVOCCCVQPTTW-----TTTTENPPTPIITTTTTPPTP 1908  
QY 718 TSTTSDKP-----APT--TPKGTAPTTPKKEPAPTTPKKEPAPTTKGTAPTLKKEPAPTK 772  
Db 1909 TPTSTQSPNGLQAPPTPISTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1967  
QY 773 PAKKELAPTTTKGPTST--TSKAPATTPKETAATTPKKEPAPTPKKP-----APTPPE 824  
Db 1968 PTTVLTITTTTTPPTTTPPTTTPSTKSTVTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 2027  
QY 825 TPTPTTSEVSTPTTTPK-----EPTTIHKSPEDESTPELSAEPKPLKLENSPKRPGPTTKT 879  
Db 2028 TPTPTTSGSPPTPHNSTAIATLITNSNPPESTTQTSISSTPLTSTTTLSTLPPAIE 2087  
QY 880 PAKTKEMTTAKDKTTERDOLRTPEETTPAARMTKETATTEKTESKITATTOVST 939  
Db 2088 MSTAPSTPTAPPTTSGHTLSPPSTTTPSPGTPRGTTT--GSSAPPTSTVQTITT 2145  
QY 940 TQDPTTPEKIT--LKTTLAPKVTYTKKITTITTEINKKPE----- 979  
Db 2146 SAMTPTPLPLSTPLSIRITGLRPYSVILICVLDITYAAGEEVYNGTYDTCYFVNC 2205  
QY 980 -----TAKPKDRAINSKATTPKPOKPTKAPKPTSTKPKTPMPVRKPKPTPT 1027  
Db 2206 LSCLEFYNMSCPSPTSPPTPK--STPTPSKPSSTPKPTPKRPPCPEPDP-- 2260  
QY 1028 PAKMTSTPELPTSRIAEMLOTTTRNONTNKLAVNP-----KSDAG 1074  
Db 2261 -----QENETWMLDCFM--ATCKNTNVEIVKVECEPPMPPTCSNGILOPVRVEDPD 2310  
QY 1075 G-----AEGTPHML-----RPHVFMEVTPDDMDYLPRVNOGITI 1111  
Db 2311 GCMHMECDYCTGMD--PHVTFPDGLYSYGNCITYLVEISPSVD-----NFGVI 2363  
QY 1112 N 1112  
Db 2364 D 2364

RESULT 2  
118262  
S-layer protein - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18262  
R:Fujino, T.; Beguin, P.; Aubert, J. P.  
J. Bacteriol. 175, 1891-1899, 1993  
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose.  
A:Reference number: Z18847; MUID:93209931  
A:Accession: T18262  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1664 <FUD>  
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 26, 2002, 16:18:10 ; Search time 82.85 Seconds  
(without alignments)  
1268.810 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_25\_1404  
Perfect score: 7410  
Sequence: 1 QDLSSCAGRCGEGYRDATC.....ARATITRSQGLSKWYNCP 1380

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR1:  
2: PIR2:  
3: PIR3:  
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1090	14.7	3020	2	A43932	mucin 2 precursor,
2	950	12.8	1664	2	T18262	S-layer protein -
3	863.5	11.7	1489	2	T31108	cyst germination s
4	846	11.4	1274	2	T16251	hypothetical prote
5	812	10.9	2187	2	T30826	nascent polypeptid
6	809	10.9	1367	1	S48478	glucan 1,4-alpha-g
7	784	10.6	1491	2	S49915	extensin-like prot
8	695	9.4	1151	2	T18835	high molecular mas
9	676.5	9.1	1229	2	T12697	hypothetical prote
10	672	9.1	1344	1	A35175	mucin 1 precursor,
11	666.5	9.0	7962	2	T38346	elastic titin - nu
12	663	8.9	3570	2	T45025	mucin MUC5B, trach
13	658.5	8.9	3507	2	T34513	hypothetical prote
14	643.5	8.7	990	2	T51618	nucleolar phosphop
15	633	8.5	489	2	T11622	extensin class 1 p
16	632	8.5	761	2	T29757	hypothetical prote
17	629	8.5	6642	2	T19431	protein UNC-89 - C
18	627.5	8.5	971	2	T75518	hypothetical prote
19	622.5	8.4	839	2	T29018	hypothetical prote
20	607.5	8.2	801	2	S27933	gene LF3 protein -
21	607.5	8.2	924	2	S50155	larval glue protei
22	605	8.2	379	2	T34434	hypothetical prote
23	591.5	8.0	2232	2	A33577	ascites stialoglyc
24	568.5	7.7	1630	2	S22456	hydroxyproline-ric
25	559.5	7.6	350	2	S23760	polyphenolic adhes
26	559	7.5	875	2	T16543	hypothetical prote
27	556.5	7.5	856	2	S06733	hydroxyproline-ric
28	551	7.4	620	2	QPM5H	neurofilament trip
29	551	7.4	1087	1	QPM5H	neurofilament trip

30	545	7.4	873	2	A47283	calphotin - fruit
31	542	7.3	369	2	S20500	hydroxyproline-ric
32	540	7.3	1459	2	T32271	hypothetical prote
33	538.5	7.3	416	2	T00465	extensin precursor
34	537	7.2	1072	1	A37221	neurofilament trip
35	532	7.2	756	2	T27642	hypothetical prote
36	530.5	7.2	813	2	S70795	vesa protein precu
37	530.5	7.2	1162	2	TJ0557	calcium-binding pr
38	530	7.2	865	2	A47282	hydroxyproline-ric
39	522	7.0	328	2	T00985	neurofilament trip
40	521.5	7.0	854	2	S02003	membrane glycoprot
41	518	7.0	866	2	T45462	hydroxyproline-ric
42	518	7.0	1611	2	T38236	tegment protein 2
43	515.5	6.9	3534	2	T42567	mucin-like glycopr
44	513	6.9	1832	2	T31113	membrane glycoprot
45	512.5	6.9	867	2	T45463	membrane glycoprot

## ALIGNMENTS

RESULT 1  
A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C:Date: 10-Mar-1993 #sequence\_revision 12-Apr-1996 #text\_change 05-Nov-1999  
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329  
R:Gum Jr., J.R., Hicks, J.W., Toribara, N.W., Siddiki, B., Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A49963; MUID:94132002  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GUI>  
A:Cross-references: GB:L21998  
R:Gum Jr., J.R., Hicks, J.W., Toribara, N.W., Rothe, E.M., Lagace, R.E., Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MUID:93016075  
A:Accession: A45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GUI>  
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GUI>  
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398  
A:Experimental source: colon  
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R:Toribara, N.W., Gum Jr., J.R., Culhane, P.J., Lagace, R.E., Hicks, J.W., Petersen,  
J. Clin. Invest. 88, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MUID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:55749; NCBIP:55750)  
R:Gum, J.R., Byrd, J.C., Hicks, J.W., Toribara, N.W., Lamport, D.T.A., Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi  
A:Reference number: A33532; MUID:89197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GUI>  
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874  
A:Experimental source: Intestine  
R:Jany, B.H., Gallup, M.W., Yan, P.S., Gum, J.R., Kim, Y.S., Basbaum, C.B.



DB 821 APPDXRPG-----STAPXAGVT-----SAPDXRPGSTAPXAGVTSAPDXRPG- 868  
QY 824 KPTKAPKKTSTKPKPTMPRVKPKTTPTRKMTSTMPELNPT-SRIAEAMLOTTTRPNQ 882  
DB 869 -GSTAPXAGVTSAPDXRPGSTAPXAGVTSAPDXRPGSTAPXAGVTSAPDXRPG 924  
QY 883 TP 884  
DB 925 RP 926

RESULT 15  
US-08-276-967-2  
; Sequence 2, Application US/08276967  
; Patent No. 5851817  
; GENERAL INFORMATION:  
; APPLICANT: Hardy, Daniel M.  
; APPLICANT: Galters, David L.  
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of  
; TITLE OF INVENTION: Sperm  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,967  
; FILING DATE: Submitted Herewith  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: UTSD:418\KIT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-787-1400  
; TELEFAX: 713-789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2476 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-276-967-2

Query Match 8.4%; Score 419.5; DB 2; Length 2476;  
Best Local Similarity 32.5%; Pred. No. 8.5e-19;  
Matches 163; Conservative 43; Mismatches 177; Indels 119; Gaps 31;

QY 406 PREPAPPTPKETAPPTPKPKLPTTPEKLAPTTPEKAPPTPEKAPPTPEEPTPTTPEEP 465  
DB 313 PSEISVSEKVPAPTE---KPTVPSIYITPTTEKPMVHMEKPIVHT--EKPTVPT-EKP 365  
QY 466 APTTPKAAPPTPKERA---PTTPKEAPPTTPKEAPPTTPKEAPPTTPKGAAPTTLKEP 521  
DB 366 TPTEKSTVPT--KKPTVFKETLPE-GPTVPAL-RPTTPEGPAVPPKG--PTVLTE- 418  
QY 522 APTTPKKAAPKLAPTTPEKSTSDKAPPTTPKGAAPT--TPKEAPPTTPKEAPPTTPK 580  
DB 419 -----WPTSHTEKSTVTEKPIPTGKSTIPTTEKPMVPTPKRI-----TTP- 458  
QY 581 GTAPPTLKEAPPTTPKKAAPKLAAPT--TTKPTSTT---SOKAPPTTPKEAPPTTPKEPA 636

DB 459 -TEKTIIPAKETV---PIEKPMVPTERTTIPTERTTIPTTEKPIVPTTEKLIAPT--EKPI 512  
QY 637 PTPPKAPAPPTPEPTPP-----TTSVSTPT-----TTKEPTTIKSPDESPTLSAEP 686  
DB 513 VPTEKPIVPTTEKHTIPTTEKLIPTTEKTIPTERTTIPTTEKPIVPTTEKPIVPT 571  
QY 687 KALENSPKEPGVPTTKTAPATKPEMTTAKDTERDRTTPEYTTAAPKMTKETATTT 746  
DB 572 -----EPTIPTTEKLIAPT--EKTIPTPKRTTPTIIRTTPTI-----RTTPT 613  
QY 747 EKTTESKITATTVQVSTTTODTTPKTIPTLKTTLIAPVTTTKTITTTTEIMNKEPETA 806  
DB 614 ERTT-----TPIIRTTPTERTT--IPKKT--VPTEKTIPT-----ERTI 652  
QY 807 KPKDRATNSKATTPKQKPTKAKPKPTSTKPKPTMPRVKPKTTPTRKMTSTMPELNPT 866  
DB 653 AP-----TTPDP-SPTLVPTOPAAVVPST-----SATTVPTPTTIASCp---FN 693  
QY 867 SRIAEAMLOTTTRPNOTPNKSL 888  
DB 694 APERCACPVSQ-SPTPNCEL 714

Search completed: April 26, 2002, 16:16:35  
Job time: 186 sec

```

1 Sequence 2 Application US/09083116
2 Patent No. 6203795
3
4 GENERAL INFORMATION:
5 APPLICANT: CHABON, Pierre
6 APPLICANT: KIENY, Marie-Paule
7 APPLICANT: LATHE, Richard
8 APPLICANT: HAREUVENI, Mera
9 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
10 TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
11 NUMBER OF SEQUENCES: 5
12 CORRESPONDENCE ADDRESS:
13 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
14 STREET: P.O. Box 1404
15 City: Alexandria
16 STATE: Virginia
17 COUNTRY: United States
18 ZIP: 22313-1404
19
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Floppy disk
22 COMPUTER: IBM PC compatible
23 OPERATING SYSTEM: PC-DOS/MS-DOS
24 SOFTWARE: Patentin Release #1.0, Version #1.30
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/09/083,116
27 FILING DATE:
28 CLASSIFICATION:
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: 08/479,537
31 FILING DATE:
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: WO PCT/FR91/00835
34 FILING DATE: 23-OCT-1991
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 08/039,320
37 FILING DATE: 04-APR-1993
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: US 08/403,576
40 FILING DATE: 14-MAR-1995
41 ATTORNEY/AGENT INFORMATION:
42 NAME: Teskin, Robin L.
43 REGISTRATION NUMBER: 35,030
44 REFERENCE/DOCKET NUMBER: 017753-025
45 TELECOMMUNICATION INFORMATION:
46 TELEPHONE: (703) 836-6620
47 TELEFAX: (703) 836-2021
48 INFORMATION FOR SEQ ID NO: 2:
49 SEQUENCE CHARACTERISTICS:
50 LENGTH: 2035 amino acids
51 TYPE: amino acid
52 STRANDEDNESS: single
53 TOPOLOGY: linear
54 MOLECULE TYPE: peptide
55 FEATURE:
56 NAME/KEY: Peptide
57 LOCATION: 128..1899
58 OTHER INFORMATION:
59 /note="The amino acids spanning
60 128 to 1899 constitute a repeated region wherein the repeat
61 20 amino acids, 17 of which are fixed. The number of such
62 OTHER INFORMATION:
63 repeats varies from 1 to 40."
64 FEATURE:
65 NAME/KEY: Peptide
66 LOCATION: 134
67 OTHER INFORMATION:
68 /note="Amino acid 134 is X1 = Xaa
69 Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,
70 CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
71 OTHER INFORMATION:
72 NAME/KEY: Peptide
73 LOCATION: 144
74 OTHER INFORMATION:
75 /note="Amino acid 144 is Y = Xaa
76 which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACZ
77 or ACG; and Asn = AAT or AAC."
78 OTHER INFORMATION:
79 FEATURE:
80 NAME/KEY: Peptide
81

```

```

: LOCATION: 147
: OTHER INFORMATION: /note= "Amino acid 147 is X2 - xaa
: OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
: OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
: FEATURE:
: NAME/KEY: Peptide
: LOCATION: 1..21
: OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
: OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match      8.5%; Score 423.5; DB: 4; Length 2035;
Best Local Similarity 24.4%; Pred. No. 3,9e19;
Matches 225; Conservative 63; Mismatches 461; Indels 203; Gaps 48;

QY 75 TSLVTKETVTEKETTNTKQSTDDKEKKTSAKETSQSIKTSKAD-LAPTSKVLAKPT 133
Db 16 TLVLT-----YSGSHASSTPGCEKETSATQSSVSPSSSEKAAVMTSSVLSHS 65

QY 134 P-KAEPTTKGR-ALTPKRP-----PTTKKEPASTTPKKEPTPTTKSAP 176
Db 66 PGGSSSTTQGDVTLAPATPRASGAATMGDVTSVPTPRALGSTTPRAHDVT---SAP 122

QY 177 TTPKEPAPTT-----KSAPTTPKRP-----APTTPKEPAPTTPKEPAPT 217
Db 123 -DNKRAPGSTAAXHAGVTSAPDXRXPGSTAAXHAGVTSAPDXRXPGSTAAXHAGVTS 180

QY 218 TKE--PAPTT-----KSAPTTKEPAPTTK---KRAPTTKEPAPTTKEPPTT 264
Db 181 APDXRXPGSTAAXHAGVTSAPDXRXPGSTAAXHAGVTSAPDXRXPGSTAAXHAGVTS 240

QY 265 PKE--PAPTTKEP-----APTTPKEPAPTAPK---KRAPTTKEPAPTTKEPAPT 311
Db 241 APDXRXPGSTAAXHAGVTSAPDXRXPGSTAAXHAGVTSAPDXRXPGSTAAXHAGVTS 300

QY 312 TKEPSTTPKKEPAPT--TKSAPTTKKEPAPT-----TKSAPTTKEPSPPTTK---E 360
Db 301 APDXRP-XPGSTAPXAHAGVTSAPDXRXPGSTAPXAHAGVTSAPDXRXPGSTAPXAHAGVT 359

QY 361 PAPTTKEPAPTTK---KRAPTTKEPAPTTK---EPAPTTKKKAPTTKKEPAPT 412
Db 360 SAPDXRXPGSTAAXHAGVTSAPDXRXPGSTAPXAHAGVTSAPDXRXPGSTAAXHAGVT 419

QY 413 TKEPAP---TTPKLTPTPEKLAPTTPKEPAPTTPELAPTTPPEEPTTPPEEAPT 468
Db 420 SAPDXRXPGSTAAXHAGVTS---APDXRXPGSTAPXAHAGVTSAPDXRXP---PGST 471

QY 469 TPKA-----AAPTTPKEPAPTTKKEPAPTTTPKKEPAPTTPEETAPTTPKG-APTTLKEPAP 523
Db 472 APXAHAGVTSAPDXRXPGSTAAXHAGVTSAPDXRP-XPGSTAPXAHAGVTSAPDXRXPXS 530

QY 524 TTP-----KKRAPKELAPTTTKEPTSTSDKRAP--TTPKGTAPTTKKEPAPTT 571
Db 531 TAPXAHAGVTSAPDXRXPGSTAP-XAHAGVTSAPDXRXPXGSTAPXAHAGVTS---APDXR 585

QY 572 KEPAATTPKG-----TAPTLTKEPAPTT-----KKRAPKELAPTTTGGPTSTSD 617
Db 586 PXPGSTAPXAHAGVTSAPDXRXPXGSTAPXAHAGVTSAPDXRXPXGSTAP-XAHAGVTSAPDX 644

QY 618 KPAP--TTPK-----ETAPTTTPKEPAPTT-----KKRAPTTPEEPTTPPEEVP 661
Db 645 RXPGSTAPXAHAGVTSAPDXRXPGSTAPXAHAGVTSAPDXRXPXS--TAPXAHAGVTSAP 702

QY 662 TTTKEPT-----IHSPPDESTPELSAEPPTPKALENSPEKEGVPPTKTPAA----- 707
Db 703 DXRXPXGSTAPXAHAGVTSAPDXRXPXGSTAPXAHAGVTSAPDXRXPXGSTAPXAHAGVTSAP 762

QY 708 -TKPEMTITAKDK--TERDLRTPEPTTAPKAMTKELATTTKEKTSKITATTTQVTS 763
Db 763 DXRXPXGSTAPXAHAGVTSAPDXRXPXGST--APXAHAGVTSAPDXRXPXGSTAPXAHAGVTS 820

764 TTQDQTPFKITLTKTTTLAPRVTTTKTTTTLTETIMNKPEETAKKRDATGATSKATTPPQ 823

```

STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22133-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/479,537A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 90/13101  
 FILING DATE: 23-OCT-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/FR91/00835  
 FILING DATE: 23-OCT-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/039,320  
 FILING DATE: 04-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/403,576  
 FILING DATE: 14-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Teskin, Robin L.  
 REGISTRATION NUMBER: 35,030  
 REFERENCE/DOCKET NUMBER: 017753-025  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2035 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 128..1899  
 OTHER INFORMATION: /note= "The amino acids spanning  
 OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat  
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
 OTHER INFORMATION: repeats varies from 1 to 40."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 134  
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,  
 OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 144  
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 147  
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 1..21  
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
 OTHER INFORMATION: 21 amino acid precursor sequence."  
 US-08-479-537A-2

Query Match 8.5%; Score 423.5; DB 2; Length 2035;  
 Best Local Similarity 24.4%; Pred. No. 3.9e-19;

Matches	235;	Conservative	63;	Mismatches	461;	Indels	203;	Gaps	48;
QY	75	TSLTVNKETTVETKETTNTTKKSTOCKETTSAKETQSTIEKTSAND-LAPTSKVLAKPP	133						
Db	16	TVLTVLTV-----VYSGHASTPGEKETTSATQSRSSVPSSTENANVSMISSVLSHS	65						
QY	134	P-KAETTKGP--ALTPKEP-----PTTKKEPASTTKPEPTTKISAP	176						
Db	66	PGSGSSSTTGQODVTLAPATPEPAGSNAATWGQDVTSVPTPALGSTPPPAHDVT---SAP	122						
QY	177	TTPEKPAPTT-----KSAPTTPEP-----APTTPKEPATTTPKEPPTT	217						
Db	123	--DNKPAAGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	180						
QY	218	TKE--PAPTT-----KSAPTTPEKPAPTTPK--KPAPTTPEKPAPTTPKEPPTT	264						
Db	181	APDXRPXPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	240						
QY	265	PKE--PAPTTKEP-----APTTPKEPAPTAPK--KPAPTTPEKPAPTTPKEPAPT	311						
Db	241	APDXRPXPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	300						
QY	312	TKEPSPTTPKEPAPT--TSAPTTTPKEPAPT--TSAPTTTPKEPAPT--E	360						
Db	301	APDXRP--XPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	359						
QY	361	PAPTTPEKPAPTTPK--KPAPTTPEKPAPTTPK--EPAPTTTPKPAPTTPKEPAPT	412						
Db	360	SAPXRPXPGSTAPXAHGVTAPDXRPXGSTAPXAHGVTAPDXRPXGSTAPXAHGVT	419						
QY	413	TPKEPAP--TPPKLTPPTPEKLAPTTPPKPAPTTPPELAPTTPEEPPTTPEEPAPT	468						
Db	420	SAPXRPXPGSTAPXAHGVT--APDXRPXPGSTAPXAHGVTAPDXRPX--PGST	471						
QY	469	TPKA-----AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT	523						
Db	472	APXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTAPDXRPXGS	530						
QY	524	TPP-----KKPAPELAPTTTPKEPTSTSDKPA--TPPKGAPTTTPKEPAPTTP	571						
Db	531	TAPXAHGVTAPDXRPXPGSTAP--XAHGVTAPDXRPXPGSTAPXAHGVT--APDXR	585						
QY	572	KEPAPTTPKG--TAPTTTPKEPAPTTP-----KKPAPELAPTTTPKPTSTSD	617						
Db	586	PXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGSTAP--XAHGVTAPDX	644						
QY	618	KPA--TPPK--ETAPTTPKEPAPTTP-----KKPAPTTPPTTPPTTSEVSTP	661						
Db	645	RPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGS--TAPXAHGVTAP	702						
QY	662	TTTPKEPTT-----IKSPDESPTPELSAETTPKALENSPEKPGVPTTKTAA-----	707						
Db	703	DXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAPDXRPXPGSTAPXAHGVTAP	762						
QY	708	-TKPEMTTAKK--TTERDLRTPEPTTAPKPKTKETATTEKTSKITATTTQVTS	763						
Db	763	DXRPXPGSTAPXAHGVTAPDXRPXPGST--APXAHGVTAPDXRPXPGSTAPXAHGVT	820						
QY	764	TTTQDTTFPKITTTKTTTLAPKVTTTKTITTTTELINMKPEETAKPKDRATNSKATTPPQ	823						
Db	821	APDXRPXG--STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTAPDXRPX--	868						
QY	824	KPTKAPKPKETSTKKKTPMRVKKPTTTPPKMTSTMELNPT--SRIAEMLQTTTPNQ	882						
Db	869	-GSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTSA-PDXRPXPGSTAPXAHGVTAPDX	924						
QY	883	TP 884							
Db	925	RP 926							

RESULT 14  
 US-09-083-116-2



FILED DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039,320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1867 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 128..1727  
OTHER INFORMATION: /note= "the amino acids spanning  
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repeat  
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 134  
OTHER INFORMATION: /note= "Amino acid 134 is X1 - xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y - xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 - xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-09-083-116-5

Query Match 8 5%; Score 423.5; DB 4; Length 1867;  
Best Local Similarity 24.4%; Pred. No. 3.5e-19;  
Matches 225; Conservative 63; Mismatches 461; Indels 203; Gaps 48;

QY 75 TSLVVKETTVETKETTNNKQTSIDGKEKETSASKESTIEKTSARD-LAPTSKVLAKPT 133  
DB 16 TVLT-----VTGSHASSTPGGEKETSATORSSVPSSTENKAVSMSSVLSHS 65  
QY 134 P-KAETTTGP--ALTTPKP-----TPTPKEPASTTPKEPTPTTKSAP 176  
DB 66 PGSGSSTTGQDVTIAPATEPAGSAATWGQDVTVPTRPALGSTTPPAHDVT--SAP 122  
QY 177 TPKEPAPTTT-----KSAPTPPKP-----APTTKEPAPTPPKPAPT 217  
DB 123 --DNKRAPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVT 180

QY 218 TKE--PAPTT-----KSAPTPKEPAPTPK-----KPAPTPKEPAPTPKEPTPT 264  
DB 181 APDXRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVT 240  
QY 265 PKE--PAPTTKEP-----APTTPKEPAPTPK-----KPAPTPKEPAPTPKEPAPT 311  
DB 241 APDXRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVT 300  
QY 312 TKEPSPPTPKPAPTT--TYSAPTTTKEPAPTT-----TKSAPTPKEPSPPTTK--E 360  
DB 301 APDXRP--XPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVT 359  
QY 361 PAPTPKEPAPTPK-----KPAPTPKEPAPTPK-----EPAPTTTKPAPTPAPKEPAPT 412  
DB 360 SAPDXRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVT 419  
QY 413 TPKEPAP-----TPPKLTPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPPTTPEEPAPT 468  
DB 420 SAPDXRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVT 471  
QY 469 TPKA-----AAPNTPKPAPTPKEPAPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 523  
DB 472 APXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGS 530  
QY 524 TTP-----KKPAPELAPTTTKEPSTSDKRAP--TTPKGTAPTPKEPAPTP 571  
DB 531 TAPXAHGVTSA PDKRPXPGSTAP--XAHGVTSA PDKRPXPGSTAPXAHGVTSA--APDXR 585  
QY 572 KEPAPTTPK-----TAPTTLKEPAPTT-----KKPAPELAPTTTKEPSTTSD 617  
DB 586 PXXGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAP--XAHGVTSA PDX 644  
QY 618 KPAP--TTPK-----ETAPTPKEPAPTP-----KKPAPTPPEPTTSEVSTP 661  
DB 645 RPYPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGS--TAPXAHGVTSA 702  
QY 662 TTTKEPPT-----TKSPDESTEPLSAEPPTKALENSPKRPGVTTTPAA-----707  
DB 703 DXRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA PDKRPXPGSTAPXAHGVTSA 762  
QY 708 -TKPEMTTAKDK-----TTERDLTTPPETTTAPKMTKETATTEKTESKITAATTOVTS 763  
DB 763 DXRPXPGSTAPXAHGVTSA PDKRPXPGS--APXAHGVTSA PDKRPXPGSTAPXAHGVT 820  
QY 764 TTTQDTTPEKITTLKTTTLPAPKVTITTKTITTEINKKEPEAKRDRATNSKATTPK 823  
DB 821 APDXRPXPG-----TAPXAHGVT-----SAPDXRPXPGSTAPXAHGVTSA PDKRPX 868  
QY 824 KPTKAPKKPTSTKKPKTTPRVRKPKTTTPPKMTSTMPELNPT--SRIAPAMLOTTTRPNO 882  
DB 869 -GSTAPXAHGVTSA PDKRP--XPGSTAPXAHGVTSA--PDKRPXPGSTAPXAHGVTSA PDX 924  
QY 883 TP 884  
DB 925 RP 926

RESULT 13  
US-08-479-537A-2  
; Sequence 2, Application US/08479537A  
; Patent No. 5861381  
; GENERAL INFORMATION:  
; APPLICANT: CHAMON, Pierre  
; APPLICANT: KIENY, Marie-Paule  
; APPLICANT: LATHY, Richard  
; APPLICANT: HAREVENT, Marc  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
; STREET: P.O. Box 1404  
; CITY: Alexandria

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1867 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 128..1727  
 OTHER INFORMATION: /note="The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat  
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
 OTHER INFORMATION: repeats varies from 1 to 40."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 134  
 OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa  
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 144  
 OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa  
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 147  
 OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa  
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 1..21  
 OTHER INFORMATION: /note="Amino acids 1 to 21 are a  
 OTHER INFORMATION: 21 amino acid precursor sequence."  
 US-08-479-537A-5

Query Match 8.5%; Score 423.5; DB 2; Length 1867;  
 Best Local Similarity 24.4%; Pred. No. 3.5e-19;  
 Matches 235; Conservative 63; Mismatches 461; Indels 203; Gaps 48;  
 QY 75 TSLVNETTEVETKETTITNKOTSDGKERTTSKETSIEKTSKAD-LAPTSKVLAKPT 133  
 DB 16 TLTIV-----VTGSGHASSTPGKEKETSATQSSVSPSSTEKNAVSWTSSLSHS 65  
 QY 134 P-KAETTKGP--ALTTPKEP-----TPTTPKEPSTPKKEPTPTTISAP 176  
 DB 66 PGSSSTTGQDVTLAPATEPASGAATGQDVTVPVTPALGSTITPPAHDTV---SAP 122  
 QY 177 TTPKEPAPTTT-----KSAPTTPKPEP-----APTTTKEPAPTTPKPEAPTT 217  
 DB 123 --DKRPAPGSTAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 180  
 QY 218 TKE--PAPTTT-----KSAPTTPKPEPAPTTK---KRAPTTPKPEPAPTTKEPPTT 264  
 DB 181 APDXRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 240  
 QY 265 PKE--PAPTTKEP-----APTTPKPEPAPTPK---KRAPTTPKPEPAPTTPKPEAPTT 311  
 DB 241 APDXRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 300  
 QY 312 TKEPSPPTPKPEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTPKPEPSPPTTK---E 360  
 DB 301 APDXRP--XPGSTAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 359  
 QY 361 PAPTTPKPEPAPTTK---KRAPTTPKPEPAPTTK---EAPTTTTPKRAPTAPEPAPT 412

DB 360 SAPDXRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 419  
 QY 413 TPKETAP-----TTPKKLTPTTPEKLAPTTPEKAPTTPEBLAPTTPEEPPTTPEPAPT 468  
 DB 420 SAPDXRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 471  
 QY 469 TPKA-----AAPNTPKPEPAPTTPEKAPTTPEKAPTTPEKAPTTPEKAPTTPEKAP 523  
 DB 472 APXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 530  
 QY 524 TTP-----KKRAPKELAPTTTKEPTSTTSOKPAP--TTPKGTAPTTTKEPAPTT 571  
 DB 531 TAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 585  
 QY 572 KEAPPTTPG-----TAPTTLKPEPAPTT-----KKRAPKELAPTTTGTPTSTSD 617  
 DB 586 PKEGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 644  
 QY 618 KRAP--TTPK---ETAPTTPKPEPAPTT-----KKRAPTTPEPTTSEVSTP 661  
 DB 645 RPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 702  
 QY 662 TTTKEPTT-----TKSPDESTPELSAPPTPKALENSPKKEGVTPTTTPAA----- 707  
 DB 703 DXRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 762  
 QY 708 --TKPEMTTAKDK---TTERDLRTPEPTTAAAPKMTKETATTEKTESKITATTTOVTS 763  
 DB 763 DXRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 820  
 QY 764 TTTQDTTPEKITLTKTTTLAPKVTITTKITTTTEINAKKEEETAKPRDRATNSKATTPKQ 823  
 DB 821 APDXRPXG-----STAPXAHGVT-----SAPDXRPXGAPXAHGVTSAPOKRPXGAPXAHGVT 868  
 QY 824 KPRKAKRPKPTSTKPKPTMPVRKPKTTPRKMTSTMPELNPT--SRIEAMLOTITRRPQ 882  
 DB 869 -GSTAPXAHGVTSAPOKRP--XPGSTAPXAHGVTSA--PDARXPGSTAPXAHGVTSAPOK 924  
 QY 883 TP 884  
 DB 925 RP 926

RESULT 12  
 US-09-083-116-5  
 ; Sequence 5, Application us/09083116  
 ; Patent No. 6203795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHAMBOU, Pierre  
 ; APPLICANT: KIRBY, Marie-Paule  
 ; APPLICANT: LARHE, Richard  
 ; APPLICANT: HAREUVENT, Mara  
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
 ; TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
 ; STREET: P.O. Box 1404  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/083,116  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/479,537

QY 798 IMNKP--EETAKPD-----RATNSKATTPKPKKPKTSNKKPKTW----- 842  
Db 1024 VYGLPVEITGLPKRDPSPDIPFNSTGTGLVDPSTGKPINNSTAGIYSGARGLPELIDENG 1083  
QY 843 -----KVRKKTTPPTPKMTS--TMPE----- 862  
Db 1084 NLEDPSTNLPIDGNQLVNENSTVSGTSTKPKPGIPVNGGVDPDEAKDQADKG 1143  
QY 863 -----LNPSTRIAEAMLOTTTRPNOTPNKSLVEVNPKSDAGAEGETHMLLRPHVEMP 917  
Db 1144 KQGLIYPPTNSINKDPVTNTOYNTTGNF-----INP--ETGKVIPLGSLNYSFENPT 1197  
QY 918 EVTPD-----MDYLPVNPNGIITNP 938  
Db 1198 QQDTEITGRKVDVTGLPYDPSTGEIIDP 1226

RESULT 10  
US-09-219-849-5  
Sequence 5, Application US/09219849  
Patent No. 6150081  
GENERAL INFORMATION:  
APPLICANT: VAN HEERDE, GEORGE V.  
APPLICANT: VAN RIJN, ALEXIS C.  
APPLICANT: BOWSTRA, JAN B.  
APPLICANT: DE WOLF, FREDERIK A.  
APPLICANT: MOORROEK, ANDREAS  
APPLICANT: WERTEN, MARC W.T.  
APPLICANT: VAN DEN BOSCH, TANJA J.  
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN  
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
FILE REFERENCE: 2728-2  
CURRENT APPLICATION NUMBER: US/09/219, 849  
CURRENT FILING DATE: 1998-12-23  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 960  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Illustrative  
US-09-219-849-5

Query Match 9.0%; Score 452; DB 4; Length 960;  
Best Local Similarity 25.6%; Pred. No 2, 8e-21;  
Matches 173; Conservative 60; Mismatches 309; Indels 134; Gaps 41;

QY 132 PTPAETTTKPPALTTTKEPTTTTKEPASTTTPKEPTTTTTSAPTTTKEPAP--TTTTS 189  
Db 72 PGPAGPGRSNDP--PGAGPAPGP--PGSRDPGPAGP--PAGPPGSRDPGPAGAGP 126  
QY 190 APTTKEPAPTTPKEPA--PTTKEPAPTTPKEPAPTTPKAPTTTKEPAPTTP--KKPA 245  
Db 127 GPPGSRDPGPAGPAGPGRSNDPGPAGPAGPGRSNDPGPAGPAGPGRSNDP 186  
QY 246 PTTTKEPAPTTP--KEPTTTTKEPA--PTTKEPAPTTPKEPA----- 284  
Db 187 P--PGAGPAGPGRSNDPGPAGPAGPGRSNDPGPAGPAGPGRSNDP--PGAGP--A 298  
QY 285 -----PTAPK--KAPTTT--KEPAPTTPKEPA--PTTKEPAPTTPKEPAPTTPKS 330  
Db 245 PAGKAGAGPAGPAGPAGPGRSNDPGPAGPAGPAGPGRSNDP--PGAGP--A 298  
QY 331 APTTKEPAPTTPKSAPTTTPKEPAPTTPKEPAPTTPKEPAPTTP--KKPAPTTPKEPAP 387  
Db 299 GPPGSRDPGP--PGAGPAGPGRSNDPGP--PGAGPAGPGRSNDP--PGAGP 348

QY 388 TTP--KEPAPTTPKPA-----PTAPKAPAPTTPKEPAPTTPKLTTPTEKLAPTT 437  
Db 349 AGPPSRDPGPAGPAGPAGPGRSNDPGPAGPAGPAGPGRSNDP--PGAGPAPGP 404  
QY 438 PEKPAPT--TPELAPTTTEETPTTTPKEPAPTTPKAA--PMTPEKPAPTTP--KEPAPT 492  
Db 405 SRDPGPAGPAGPAGPGRSNDP--PGAGPAGPGRSNDPGPAGPAGPAGPGRSNDP 462  
QY 493 TPEKPA-----PTTPEK--APTTPKGT-----APTTLKAPAPTTPK--APKELA 535  
Db 463 GAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 532  
QY 536 PTTTKEPTSTSDKAPAPTTPKGT--PTTKEPAPTTP--KEPAPTTPKGT--PTTLEK 590  
Db 523 PPGSRDPGPAG--PGAGPGRSNDPGPAGPAGPGRSNDPGPAGPAGPGRSNDP 580  
QY 591 APTTTPKRPAP-----KELAPTTTTPKGTST--TSDKAPAPTTPKETAPTTTPKEPAPTTPK 642  
Db 581 GP--PGAGPAGPGRSNDPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 636  
QY 643 PAPTP-----ETPPTTSEVSTPTTTPKEPTTTTTPKEPTTTPKEPTTTPKEPTTTPKEPTTTP 693  
Db 637 PGPAGPGRSNDPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 693  
QY 694 PKEGPVPTTTPATK 709  
Db 696 SRDPGPAGPAGPAGP 711

RESULT 11  
US-08-479-537A-5  
Sequence 5, Application US/08479537A  
Patent No. 5861381  
GENERAL INFORMATION:  
APPLICANT: CHAMON, Pierre  
APPLICANT: KIENY, Marie-Paulle  
APPLICANT: LATHE, Richard  
APPLICANT: HAREJVENT, Mara  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SNECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479, 537A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 90/13101  
FILING DATE: 23-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00835  
FILING DATE: 23-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/039, 320  
FILING DATE: 04-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403, 576  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Teskin, Robin L.  
REGISTRATION NUMBER: 35,030  
REFERENCE/DOCKET NUMBER: 017753-025

```

Db 1133 DEAKAGQADKGNKGLVLPNTNSINKDPVTNTQYSNFTGNINPDKVLPGLPSLNP 1192
QY 675 DESTPELSAEPYPPKALE---NSKEGCVPTTKPAATKPEMTTAAKDKTERDLRTTPEP 731
Db 1193 SFMTPOOTDEITKRPDYVTLGLYDPSGTEIIDPAKRLPGSVADDELITTEVLTITDE 1252
QY 732 TTAPAKMTKETATTEKTESKTTATTTQVSTTTQDTTPPKITTLTKTLAKVTTTK 791
Db 1253 VTGLP-IDLETGLPDPD-----VSGLPOLPNTLVD-----PSNKK 1287
QY 792 TI-----TTTEIMNKPEETAKPD-----RAFNK 816
Db 1288 PIGSHSGFINGTSGEQSHKDPSTGKPLDPNTGLHPFDEDSGLINPBTGDKLQSGSHG 1347
QY 817 AITPKPKP-----TKAPKPTSTKKPKTMPVRKPKTTPPKMTST-----859
Db 1348 TFMFVPGKPGENGIMTPPOILEALKLPSTSEVNISPR---PSSDAVDRPTNTMMNK 1404
QY 860 -----MPLNPTSHIAEAMLOTTTRPNQ-----PNSKL-----VEYNPKS 895
Db 1405 ISGOTYVDKKTILGSAASVITHALCTPTQDTPTGLPSDPSGTLPIFGFNVLDPPQ 1464
QY 896 EDAGAGETPHMLLRPHVEMPEVTPMDV-LPRVNOGIIINPM 939
Db 1465 GEQ--IKGSVYVSL--YVKEKNIVTEAAYGLPVPKTPGPIDPT 1505

```

RESULT 9  
US-08-928-361B-6  
Sequence 6, Application US/08928361B  
Patent No. 6071518

## GENERAL INFORMATION:

APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
TITLE OF INVENTION: SPECIES INFECTIONS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PETERS, CARYN, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto

COUNTRY: USA

ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08-928, 361B

FILING DATE: 12-SEP-1997

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 60/026,062

FILING DATE: 13-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: VERNY, Hana

REGISTRATION NUMBER: 30,518

REFERENCE/DOCKET NUMBER: 480,76-1(HV)

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-1677

TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1721 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-928-361B-6

Query Match 9.28; Score 463.5; DB 3; Length 1721;  
Best Local Similarity 22.18; Pred. No. 9.7e-22;  
Matches 271; Conservative 78; Mismatches 467; Indels 413; Gaps 39;

```

QY 5 KKNRFRKKRTPKPPVADAGSGLDNGDKVTPDSTTQHNKVSSTPKITTAKPINRPS 64
Db 116 KSNRKK---TTPPSANTAGVRSN-ETKTEPSANT---NFLVADPKI-----N 158
QY 65 LPPNSDTSKE-----TSLVNKETTVEKETTITTKKOSTGCKEKTTSK 109
Db 159 ACONSSESEGOQIFDMGSKVYIPYTKCVGKHRTTTTTTTTTTTTTTTTTTTT 208
QY 110 EQSISTKTAQKALAPTSKVLAKPTPKAETTTGPGALTTTPKEPTTPKEPASTPKPEPTP 169
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 170 TTIKSAPTPKPEAPPTTKSAPTPKPEAPPTTKPEAPPTTKPEAPPTTKSA 229
Db 241 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300
QY 230 PTPKPEAPPTPKKPAATTPKPEAPPTPKPEPTTPKPEAPPTKEAPPTKEAPPTAPK 289
Db 301 TTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 350
QY 290 KPAETPKPEAPPTPKPEAPPTTKPEPSPTPKPEAPPTTKSAPTTKPEAPPTKSAPT 349
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTATTTT 406
QY 350 PKR-----PSPTT-----357
Db 407 TSETESVIRKPEDEMCKLEKNGCEAKGATYGVIGKDGRIENGMAFTMINDDTHVFRK 466
QY 358 -----TKPE-----APTTPKE 368
Db 467 VKDQNTISVRCRKGAGKLEFPDRSLDFTTPPAAGHNSCSIIYVSGDCKIHSPYSGSD 526
QY 369 PAPTTPKPAETTPKE-----PAPTTPKPEAPPTTKKPAATKAPK 410
Db 527 VSLS-----APIQSELFNEVYCDCTAKYGAHSGYQISADVTTTAKPTTT--TGA 580
QY 411 PTPKPEAPPTPKKLPPTPEKLAPTTPKPEAPPTPEELAPPTPEEPTTPPEAPPTP 470
Db 581 PGQPTTTTGSPSKPTTTTAKATTT-----TLLNPITTTTOKPTTT-----TT- 627
QY 471 KAAPNTPKPEAPPTPKPEAPPTPKPEAPPTPKETAAPTTPKGAAPTTLKPEAPPTPKPA 530
Db 628 --KVPKRPATTTTTLKPITTTTTKATTTTTTTPVT-----TTTKRDEMTTTTTTL 680
QY 531 PK-----ELAPTTKEPTS-----TSDKPAPTTPKG-TAPTTPKEAPPT 570
Db 681 PDIGIEITPIPIKMDKYRMITYDYNSSLGILDSNDEPIPGSAGQIADISLNFQYTH 740
QY 571 PKPEAPPTPKGTAPTTLKE---PAPTTPKPAPELAPPTTKGPTSTYSDKPAPTPKET 627
Db 741 KSTGLPIDPMGLPDRKSGMLVHPYTNQMSGLSVSLAAKNLTVPTDEYTG--LPIDT 798
QY 628 APTTPKEAPPTPKKPAETTPETPTTSEVSTPT-----TKPEPTIHKS-----673
Db 799 LTGYLDVSLIPFN--PBTGELFDPLISDEIMNIGIYSGISASISLSSOKALIDPA 856
QY 674 -----PDEST-----BELSAEPPKALENSPK 695
Db 857 TNNVGERGGLINPATGMIPEGLPSQOTOPSEIHDGGIILPEVAANADKFKLSIP- 915
QY 696 EPGVPTTKPATKPEMTTAAKDKTER-----DLRTTETTTAPAKM 738
Db 916 -PSVP-----ESIDE-----KKQKIDSISELMYDIESGRILGOVSKRPIDGSIAGDLP 963
QY 739 TKETATTEKTESKITAT--TQVSTTTQDTTPPKITLTKTLTTLAPVTTTKTITTTTE 797
Db 964 IMKTPTQTDSTVYKPIDPTTGLPFPNPTGHLINPNNNTMOSSPAGAKYAVANSIKIDN 1023

```

```

Db      11 PKMTYPPYKAKSPYPP-----TYKSKPTY-----KPKIT----- 40
Qy      111 TOSIEKTSANDALSTSVLAKPTPKAEETTCKGALPTPKKEPPTPKKE-----PASTPKKE 166
Db      41 -----YPTPK-----AKPS-----YPTPKPKKTTYPTPKKLYPTPKPKP 79
Qy      167 PTPPTTISAPL-TPKEPAPPTTTSAPPTPKAPPTTKEPAPPTTKEPAPPTTKEPAP--P 223
Db      80 SYPTPKSYKPTPKKTYKTYPTPKAKPSYPTPKPKKTYPTT-YKPLTYTPPKPKKASYP 138
Qy      224 TTTTSAPPTTKEPAPPTPKKAPPT-TPKEAPAPT-----KPEPTTPKEPAPPTTKEPAPT 278
Db      139 PTKPKKPSYP--PSYKTKKTYPTPKKLYPTPKPSYPSYKPNKTYPTPKPKLT 196
Qy      279 TPKEPAPTAAPKCA--PTTPKEBA--PTTPKEPAPPTTKEPSPTPP-----KEPAPTTK 329
Db      197 YP-----PTYKAKPSYPTPKAKSPYPTKAKSPYPTKAKPSYPSLAKAPSPYPTK 252
Qy      330 SAPPTTKEPA-----PTTTSAPPTPKPEPSPITTTKEBA--PTTPKEBA-----PTTPKK 376
Db      253 AKPTPKAKPTPKKAPPTPKPSYPTKAKPYPP---PTYKAKSPYPTPKAKPTPKAKPTPKAK 308
Qy      377 PA--PTTPKEPA--PTTPKEPAPPTTTPKAPAPKKEPAPPTTKEPAPTTPPKKLYPTTPEK 432
Db      309 PSYPTPKAKPSYPTPKAKSPYPTKAKSPYPTKAKSPY-----PTYKAKSPYPTPKKPT--YK 362
Qy      433 LATTPPE--KPAPTPEELAPTTPEEPT--PTTPEEPA--PTTPKAADNTPKEPAPTTPK 487
Db      363 AKPTYSITTKAKPSYP---PTYKAKSPYPTPKAKSPYPTPK--AKPSYP---PTYKA 413
Qy      488 EPA-PTT---KPEAPPTPKETAPTT--PKGTAPTTTKEPAPTTPKPKAPKELAPTTK 540
Db      414 KPTPKSYTKAKSPYPTPKKISTYPTPKAKPSYSTTKAKSSYPTPKAKPSYPTPKA 473
Qy      541 EPTSTSDKPAPTTKGAPPTPKAPAPTTPEKAPPTPKGTAPTTTKEBA--PTTEPKP 598
Db      474 KPTPKSYTKAKPTPK--AKPSYP---PTYKAKPS-----YPTPKAKPSYPTPKAKP 521
Qy      599 ABEKELAPTTTKGPTSTSDKPAPTTPKETAPTTPEKPA--PTTTPKKA--PTT---PETP 651
Db      522 TYK-AKPTPKAKPSYPTPKAKPSYP-----PTYKAKSPYPTPKAKPYPSYPTPKAKSPY 576
Qy      653 PPTSEVSTPTTTPKPTTTHKSPDESSTPELSAEPTPKALENSPKEGCVPTTKTPAATPKPE 711
Db      577 PTKPKKPSYPTP-----YKAKPSYPTPKAKPTKA-----KPTPKSP--YKAKPS 620
Qy      712 MTTTAKDTERDLATTPPETTTAAP 736
Db      621 YPTPKAKPTPKAKPSYPTPKAKP 645

RESULT      8
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLIN
; APPLICANT: LEBCH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUY, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; NUMBER OF SEQ. ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ. ID NO. 5
LENGTH: 1721
TYPE: PRT

```

ORGANISM: *Cryptosporidium parvum*  
US-08-700-651-5

Query Match	9.3%;	Score 466.5;	DB 3;	Length 1721;
Best Local Similarity	20.9%;	Pred. No. 6.3e-22;		
Matches 285;	Conservative 100;	Mismatches 473;	Indels 507;	Gaps 45

Qy	33	KVTPDSTVOHKNVSTSPKITTAKPIINRPSPJSPMSDSKETSLSLVNKEVTEKNTT	92
Db	190	KHNTTT	24
Qy	93	TNKOTSDGKEKTSKETSISIKTSAKDAPSKVLAKPTKAEITTTKGPAITTPKEPT	15
Db	250	TT	30
Qy	153	PTPKPEASTTPKEPTPTTITKASAPTTPKPEAPPTTKSAPPTPKPEAPTTTKEPATTPKE	21
Db	310	TT	36
Qy	213	PAPTTPKEAPTTTKSAPTTPKPEAPPTPKKAPPAPTPKPEAPPTPKPEPTTPKE	26
Db	370	TTTTTTTTTKPTTTTTTTTTTT-----TTTTKKPTTTTTATTTTTTSESESVKIPBEMCLE	42
Qy	268	-----PAPTTPK-----	27
Db	424	KNECEAKGATYGVIGKDGRIENGKAFMTIPIINDTJHVRFRKXYKDVGNTISVRCRKAG	48
Qy	274	-----EPAPTTPKEPA--PTPKKPA-----	29
Db	484	KLEFPDRSIDFTPIPVAGHNSCSIIIVGSGDGKINVSRYSGKDVLSIAPJOPSELENEV	54
Qy	293	-----PTPKPEAPTT--PKEAPPTTKEPS--PTPKPEKAP	32
Db	544	YCDTCAKIGALHSGYOSADSVTTTTTAKPTTTTTTGAGGCPPTTTTTSPSKPPTTTTTTKA	60
Qy	326	TTTTKSA-----PTTKPEAPTTTKSAPTTPKEPSPTTPKPEAPTTPKPEAPTTPKKAP	38
Db	604	TTTTTTTTLPIITTTQKPTTTTTTTVBPCKPFIATTTTTLTKIYTTTTTKATTTTTTTTTVPT	66
Qy	381	TPKRAPPTPKKAPPAITTKKPA-----	40
Db	664	T-----TTTTKHEMTTTTTPLPIDIGIEIPIPIEKMLDKYTRMIYDVSGLLSDNDE	71
Qy	409	PAP-----PTPKETA--PTTP-----	42
Db	719	PIPSGQAGIADTSLNLPFVQIHKSTGLPIDBMGGLPRDPKSGMLVHPYTNQTMGSLSVSY	77
Qy	423	---KKLT-----PTPKELIAPTPE-----	43
Db	779	IAAKMLIVDDETYGLPIDTLGVPRLDPSVLIPIPNPBTGLFLPDISDELINGTIACIVSG	83
Qy	440	-----KPAPTTPBE-----LAPT-----PEBPTPTP--EBAP	46
Db	839	ISASSLSLSOKSALIDPTNNVGEFGGLINPATGVMIPEGLPSBOGOFSPEDGII	89
Qy	469	TPKAAAPTTPKEAPTTPKPEAPTTPK-----	49
Db	899	PREVAANAADKFKLISIPSPVESIPEKDOXIDISIELMWDIESGRILGVGSKRPIGSA	95
Qy	497	-----PAPTTPKETAPTTPKGTAPTTLKEAPPTPKKAPKAPKELAPT--TTKEPTSTSDKP	55
Db	959	GDLNPIKMTPTQDTSVGKPIDPTT--GLPFPN--PTGHLINPTNNNNMDSFPAKATKY	101
Qy	551	APT-----PKGTAPTTPKEPA-----PTPKPEAPPTPK-----	58
Db	1014	AVSNGICITDNVYGLPVGELTGLPKRDSODIPFNSITGELVJDSTGKPIINNSTAGIVSGKP	107
Qy	581	GTAAPT-----TLKEAPTTPKKAPKAPKELAP--TTTKGPTSTSDKPAPTTPKETAAPT	63
Db	1074	GLPIEDENGNLFPDSNLPIDGNNGLVNPENSTYSGTSGTT--KPKGCLPIVNGGVAVP	113
Qy	633	KEPA-----PTPKKAPAPTPEPTTPPTTSEVSTPTTTKE--PTTIHKS--P	67

```

OY 335 TKEPAPTTKSAPTTPKESPTTKEPAPTTPKEPA-PTTPKKP-APTTPKEPAPTTPKE 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 444 PKKPMP-----NPSNPNKPN-----EPINPESNPNPSPNAPSNE--PSNPN 492
OY 393 PAPTTTKKAPAPAKPEP-APTTPKETAPTTPKKLPTTTEKLAPTTPKEKAPTTPKEELAP 451
    PS-----NPNPSPNPNPSPNPN--PSNPKK-----PSNPN--P 523
DB 493 PS-----NPNPSPNPNPSPNPN--PSNPKK-----PSNPN--P 523
OY 452 TTPPEPTTPPEEPAPTTPKAAAPTTPKEP-APTTPKEPAPTTPKEPAPTTPKEAPTTP 510
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 524 SNPN--PLNPNEP-----SNPNPSPNPNPSPNPN--PSNPKP--PSNPN----- 564
OY 511 KGTAPTTLEKAPAPTTPKKAPKELAPTTPKEPTSDKAPAPTTPKGTAPTTPKEPAPTTP 570
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 565 -----PSNPEEPNPEE--PSNPKP-----SNPEEPINPEELNPKESNPNESN 606
OY 571 PKPAPPTPKGTAPTTLKEPAPTTPKKAPAKELAPTTPKGTSTSDKAPAPTTPKETAPT 630
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 607 PKKPI-----NPNPSPNPNPSPNPN--INPEDNENPLIODEPIEPNDSNVPI 649
OY 631 TPKEP-----APTTPKKAPAPTTPETPTTSVSTPTTKEPTTIHKSPESTPELSAE 684
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 650 LPIIPKGNIPSNLPENPSDSEVEYPRPNDNENSNNTKSKKNT---PNEPIPSPGDN 706
OY 685 PTPKALENSPK-----EPGVPTTK 703
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 707 PYKGHEERIKPHRSNDYVYDNNVKNKNKDEPEIPNNE 745

RESULT 6
PCT-US92-00018-2
: Sequence 2, Application PC/TUS9200018
: GENERAL INFORMATION:
: APPLICANT: Hoffman, Stephen L.
: APPLICANT: Chaoenvit, Yupin
: APPLICANT: Hedstrom, Richard
: APPLICANT: Khushf, Srisin
: APPLICANT: Rogers IV, William O.
: TITLE OF INVENTION: Protective malaria sporozoite surface protein
: TITLE OF INVENTION: Immunogen and gene encoding
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: A. David Spevack
: STREET: NMDC Building 1 T-12 National Naval
: STREET: Medical Center
: CITY: Bethesda
: STATE: MD
: COUNTRY: USA
: ZIP: 20814-5044
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/00018
: FILING DATE: 19920103
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Spevack, Avram D.
: TELEPHONE: (301) 295-6759
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 295-4033
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 826 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US92-00018-2

```

```

Query Match 9.7%; Score 488.5; DB 5; Length 826;
Best Local Similarity 26.4%; Pred. No. 1.2e-23;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

OY 168 TPTTKASAPTTPKPA-PTTTKASAPTTPKPA-PTTTPK-----APTTPKEP-APTTPKE 220
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 273 TPCKVRDQCPQIPIPIVINKIPEKPSNPEEYVNDNDPNNPNPNPNPNPNPN 332
OY 221 PAPTTTKSAPTTPKEP-APTTPKKP-APTTPKEP-APTTPKEP-TPTPKEPAPTTPKEPA 276
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 392
OY 277 PTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPSPTTPKEPAPTTPKSAPTT 334
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 393 RNPKRNPKNKPKPKPNKPNKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
OY 335 TKEPAPTTKSAPTTPKESPTTKEPAPTTPKEPA-PTTPKKP-APTTPKEPAPTTPKE 392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 444 PKKPMP-----NPSNPNKPNPN-----EPINPESNPNPSPNAPSNE--PSNPN 492
OY 493 PS-----NPNPSPNPNPSPNPN--PSNPKK-----PSNPN--P 523
DB 493 PS-----NPNPSPNPNPSPNPN--PSNPKK-----PSNPN--P 523
OY 452 TTPPEPTTPPEEPAPTTPKAAAPTTPKEP-APTTPKEPAPTTPKEPAPTTPKEAPTTP 510
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 524 SNPN--PLNPNEP-----SNPNPSPNPNPSPNPN--PSNPKP--PSNPN----- 564
OY 511 KGTAPTTLEKAPAPTTPKKAPKELAPTTPKEPTSDKAPAPTTPKGTAPTTPKEPAPTTP 570
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 565 -----PSNPEEPNPEE--PSNPKP-----SNPEEPINPEELNPKESNPNESN 606
OY 571 PKPAPPTPKGTAPTTLKEPAPTTPKKAPAKELAPTTPKGTSTSDKAPAPTTPKETAPT 630
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 607 PKKPI-----NPNPSPNPNPSPNPN--INPEDNENPLIODEPIEPNDSNVPI 649
OY 631 TPKEP-----APTTPKKAPAPTTPETPTTSVSTPTTKEPTTIHKSPESTPELSAE 684
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 650 LPIIPKGNIPSNLPENPSDSEVEYPRPNDNENSNNTKSKKNT---PNEPIPSPGDN 706
OY 685 PTPKALENSPK-----EPGVPTTK 703
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 707 PYKGHEERIKPHRSNDYVYDNNVKNKNKDEPEIPNNE 745

RESULT 7
5202236-13
: Patent No. 5202236
: APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
: SUSAN L.; MCCANDLISS, RUSS; WET, JENA; FILIPULA, DAVID
: TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
: PROTEIN
: NUMBER OF SEQUENCES: 39
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/528,762
: FILING DATE: 25-MAY-1990
: APPLICATION NUMBER: 82,456
: FILING DATE: 07-AUG-1987
: APPLICATION NUMBER: 933,945
: FILING DATE: 24-NOV-1986
: APPLICATION NUMBER: 650,128
: FILING DATE: 13-SEP-1984
: SEQ ID NO:13:
: LENGTH: 652
: 5202236-13

Query Match 9.5%; Score 476.5; DB 6; Length 652;
Best Local Similarity 29.7%; Pred. No. 5.3e-23;
Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;

OY 51 PRTTAKPINRPSLPPNSDTSKETSITLVNKEETVETKETTITTKKQISTDKKETTSAKE 110
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

OY	33	KVTPROSTONHKWVSFSPKITTAKILNRPISLPNSOSTSKETSJLVNKETTVETKETTTT	92
Dd	307	KHITTT	368
OY	93	TNKOISTDGEKEKTSKAKETOSIEKTSAKDAPLPSKVLAKPRTKATTTTKGALITPKEPT	15
Dd	367	TTTTTTTTTKKPRTT	426
OY	153	PTRPKREASTTRKEPRTTIKSAPTRPKER--APTTRKSAPTRKEBAPTTRKEBAPTTR	210
Dd	427	TTT	488
OY	211	KEBPATTTTKEBPATTTKSAPTTRKEBAPTTRKKAPTRPKREAPTTRKE-----	258
Dd	487	TTTATTTTKKPRTTTTTTTTTTTTTTTKKPRTTTTTAFTTTTTSETESVIKDEMCMLKECE	546
OY	260	-----PRTT-----PK	266
Dd	547	AKGATVGVYGKDGRIENGMATPIRNDITHRFKFKVDGNINISRCGKGAKLEFPD	606
OY	267	EPAPTTRKEP-----APTTRKEBA-PTAKKP-----	291
Dd	607	RSIDFTIPRYAGHNSCSIIIVSGGGKIHVSPYSKDVSLISAIOCPCELFENEYCDTC	666
OY	292	-----APTTRKEBAPT---PKREBAPTTRKEPS-PTTRKEBAPTTRKS-330	726
Dd	667	AKYGAIHSGYOTSADPVNTTAKKPTTTTTTGAPRGQPTTTTTGSPEKPRTTTTTYKAATTTTT	726
OY	331	----ATTTRKEBAPTTSASPTTRKEPSPTTRKEBAPTTRKEBAPTTRPKKAPTRTRKEBA	386
Dd	727	LNPITTTTTOKPRTTTTTTTKVKBEKPIATTTTTLKPIVTTTTTKATTTTTTTTTVT	781
OY	387	PTTRKEBAPTTRTKKAP-----TAREBAP-----	411
Dd	782	TTTTKRDEMTTTPRLPDIGIDEITPRIEKMLDKYRMAYDNISGLSDSDEPIDGSO	841
OY	412	-----TPKETA-PTTP-----KKI	425
Dd	842	ACOIADTSLNRPVQHNKSTGLRIDPMVGLEFPDKSGNLVHPRTNQMGSLSVYLAAKNL	901
OY	426	T-----PTTRKEKLAPTRPE-----	439
Dd	902	TVDDETGYGLPIDTLGLRIDPVLSLIRPNBETGELDPDISDEIMNGTIAIVSGIASSES	961
OY	440	----KRAPTRPEE-----LAPTT-----PEEPRTTP--EBAPATTPRKAA	474
Dd	962	LISOCSAEIDPATNMVVGEFGGLNPAITGMIRGSLGPSQOTPFSEIEDGCIIPPEVA	1022
OY	475	PNTPREBAPTTRKEBAPTTRKE-----PA	498
Dd	1022	ANADFCKLISPVSIESIPERDKOIDSISELMADIESGRILGOVSKRPIGSIADGLMPI	1081
OY	499	PTTRKEBAPTTRPKGTABPTTRKEBAPTTRPKKAPKELAPTTTKEPTSTSIDKAPPTTNGT	558
Dd	1082	MKTPQTODSVYCGKIDPTT---GLFRNP--PTGHLLNPNNMTOSSSEFAGAVKXVASNGI	1136
OY	559	APT---PKREBAPTTRKEBAPTTRPGTAPTLKEBAPTTP-----KKRPAK--	601
Dd	1137	KTDNVYGLPVEITGLPKDEVSDIPFNSTTGLVDSOCKP-INNYTAGIVSKRCLPRIE	1196
OY	602	----ELAPTTTKG-----NST-----SOKRAPTTTRKEBAPTTRKEBA---	636
Dd	1197	DENGNLDPSTKCLPIDGNNOVLVPERTNINSVSGTSGSTIKPKGPIVNGCGVAPDEEADQ	1256
OY	637	-----PTTRKKAPATTPERTPPPTTSEVSPTTTTKE-PTTIHKS---PODESREL	681
Dd	1257	ADKGDGLIVPPTINSINKOPVTNTQYSNNTGNINIBERGKIVPSGLPSGLPSENTPQQ	1316
OY	682	SABEPTKALE---NSPKREGVPPTTKPRATRKEMNTTTAKDKTTEBDLTPTPETTTAARKM	738
Dd	1317	TDELTKRPVTVYCVGLPYDSTGEIETIDPAKKLIPGSVAGDELTELVENLITTTDEVYTGJR-I	1375

RESULT 5  
US-07-638-431-2  
; Sequence 2, Application US/07638431  
; Patent No 5198535

GENERAL INFORMATION:

APPLICANT: Hoffman, Stephen L.  
APPLICANT: Charbonnet, Valerie

```

AFFILIANT: yupin
CHILDRENT: Richard
APPLICANT: Hedstrom, Richard

```

APPLICANT: khusmith, srising

APPLICANT: Rogers IV, William

Protective

```

; TITLE OF INVENTION: Immunogen
; NUMBER OF SEQUENCES: 2

```

CORRESPONDENCE ADDRESS:

ADDRESSEE: A. David Spevack

STREET: NMRDC Building 1 T-1  
STREET: Medical Center

SINCE: Medical Center  
CITY: Bethesda

STATE: MD

COUNTRY: USA

21P: 20814-5044  
COMPUTER READABLE FORM.

```
COMFOIER READABLE FORM:
MEDIUM TYPE: floppy disk
```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: IIS/07/63

FILING DATE: 19910110

CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:  
: NAME: Scott A. ...  
:NAME: SPEWACK, AVIOM D.  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 295-6759

TELEFAX: (301) 295-4033

```

; INFORMATION FOR SEQ ID NO: 2:
SEQUENCE AND FEATURE

```

SEQUENCE CHARACTERISTICS:  
LENGTH: 826 amino acids

TYPE: AMINO ACID

TOPOLGY: linear

MOLECULE TYPE: protein

05-07-638-431-2

Query match 9.78; Score 100

Best Local Similarity 26.4%; Pre Matched 153

malchies 153; conservative 54;

168 ТРТИКСАПТТПКЕРА-ПТТТКСАПТТ

[illegible]

DB 273 TPCKVRDCPQIPIPVIPPNKIPEKPSN

221 PAPT<sup>TTKS</sup>APT<sup>TP</sup>KER-APT<sup>TP</sup>KRB-A

— : —

Db 333 PNNPNNPNNPNNPNNPNNPNN

277 РТТБКЕРДРТАНУ - КД, РТТБКЕРДРТАНУ - КД

21 211 FIFTEEN APIARYS -- KRAPIVITSKIY PRIBOR

393 RRNPKRNPKNKPKPNKPNNP

GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
APPLICANT: Wang, Ping  
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
TITLE OF INVENTION: CDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pimnisi & Michaels, P.C.  
STREET: 118 No. 6187558th Tioaga  
City: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,423A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: Ptl-39  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 786 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-423A-3

Query Match	9.9%	Score 498.5	DB 4	Length 786
Best Local Similarity	29.1%	Pred. No. 2.6e-24		
Matches 200	Conservative 25	Mismatches 207	Indels 255	Gaps 32
QY 132	PTKAEETTKGALPTPKPEPTTPKPEBSATPKPEPTTLKSAPTPKPEAPPTTKSA	191		
Db 93	PGCPAAE-TTQAPATTT---QAPPTTQAPPTTT-----QAPTTTTQ--ATTITQAA	135		
QY 192	TTPKPEAPPTTKPEAPTTKEPEAPPTTKPEAPTTTKSAPTPKPEAPTTPKPA	251		
Db 136	TTTTQ--APPTTQ--ATTT---QAPTTTTQ--APTTTTQAPTTTTQ--APTTTQ--APTTTQ--	183		
QY 252	PAPTPKEPTPTTPPKPEAPATTKPEAPTTKEPEAPTPAPKPAAPTTPKPEAPT--PKPA--	308		
Db 184	-APTTQAPPTTQAAATTPAATTPAATTPAATTPAA--TTPATTPGVPAPASAPVMPDPC	241		
QY 309	-----PPTTKPESPPTPK--	321		
Db 242	ELLNGCRADFDIHLPLPHDKYCNLPFYQCSNGYTFEQRCPGELFENFPVQDCSDPANVEC	301		
QY 322	---EPAPTTT---	328		
Db 302	DGEISPAVPATYEGNEDEDIDIGDLLDNGCPANFEIDMLPHGNCRCDKYQCVHGNLYERR	361		
QY 329	-----KSAPTTKPEAPTTTKSAPTPKPESPPTTKPE	361		
Db 362	CGAGTHSEFFELQOCDAHIELVGCTLPGESESEVDDEDA--CTGWCYPTPEIEMEPILPNCF	420		
QY 362	APTT--PKP-----APTPKPKAPTT-----	381		

[illegible]

```

US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-5

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:29 ; Search time 49.78 seconds  
(without alignments)  
425.384 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1140  
Perfect score: 5011  
Sequence: 1 VKDNKKNTKKKPKPPV.....DMDYLPVNOGIINPMLS 941

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 212252 segs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	522	10.4	805	4	US-09-103-429A-4
2	506.5	10.1	744	6	Sequence 4, Appli Patent No. 5202236
3	498.5	9.9	786	4	Sequence 3, Appli Patent No. 5202236
4	493	9.8	1837	3	Sequence 5, Appli Patent No. 5202236
5	488.5	9.7	826	1	Sequence 2, Appli Patent No. 5202236
6	488.5	9.7	826	1	Sequence 2, Appli Patent No. 5202236
7	476.5	9.5	652	6	Sequence 5, Appli Patent No. 5202236
8	466.5	9.3	1721	3	Sequence 6, Appli Patent No. 5202236
9	463.5	9.2	960	4	Sequence 5, Appli Patent No. 5202236
10	452	9.0	1867	2	Sequence 5, Appli Patent No. 5202236
11	423.5	8.5	1867	2	Sequence 5, Appli Patent No. 5202236
12	423.5	8.5	1867	2	Sequence 5, Appli Patent No. 5202236
13	423.5	8.5	2035	2	Sequence 5, Appli Patent No. 5202236
14	423.5	8.5	2035	2	Sequence 5, Appli Patent No. 5202236
15	419.5	8.4	1185	4	Sequence 2, Appli Patent No. 5202236
16	419	8.4	1185	4	Sequence 2, Appli Patent No. 5202236
17	417	8.3	829	1	Sequence 5, Appli Patent No. 5202236
18	417	8.3	829	1	Sequence 5, Appli Patent No. 5202236
19	417	8.3	837	1	Sequence 5, Appli Patent No. 5202236
20	417	8.3	837	1	Sequence 5, Appli Patent No. 5202236
21	417	8.3	837	1	Sequence 5, Appli Patent No. 5202236
22	417	8.3	837	1	Sequence 5, Appli Patent No. 5202236
23	417	8.3	837	1	Sequence 5, Appli Patent No. 5202236
24	417	8.3	837	1	Sequence 5, Appli Patent No. 5202236
25	413.5	8.3	907	3	Sequence 5, Appli Patent No. 5202236
26	413.5	8.3	907	3	Sequence 5, Appli Patent No. 5202236
27	404.5	8.1	408	1	Sequence 5, Appli Patent No. 5202236

28	404.5	8.1	408	4	US-08-475-411A-65	Sequence 65, Appli
29	404.5	8.1	408	4	US-08-478-029A-65	Sequence 65, Appli
30	398.5	8.0	682	1	US-08-642-255-126	Sequence 126, Appli
31	398.5	8.0	682	1	US-08-397-633A-36	Sequence 36, Appli
32	381	7.6	1848	4	US-08-296-791-6	Sequence 6, Appli
33	381	7.6	1848	4	PCT-US95-10661A-6	Sequence 6, Appli
34	376	7.5	960	4	US-09-103-429A-4	Sequence 2, Appli
35	374.5	7.5	1537	1	US-08-325-267A-2	Sequence 84, Appli
36	371	7.4	761	2	US-08-642-255-114	Sequence 114, Appli
37	371	7.4	762	1	US-08-642-255-114	Sequence 26, Appli
38	371	7.4	762	1	US-08-397-633A-26	Sequence 62, Appli
39	370.5	7.4	1064	1	US-08-642-255-62	Sequence 4, Appli
40	369.5	7.4	1231	3	US-08-904-263A-4	Sequence 120, Appli
41	368	7.3	762	1	US-08-642-255-120	Sequence 31, Appli
42	367.5	7.3	762	1	US-08-397-633A-31	Sequence 28, Appli
43	367.5	7.3	1187	1	US-08-320-559-28	Sequence 28, Appli
44	367.5	7.3	1187	1	US-08-545-860D-28	Sequence 28, Appli
45	367.5	7.3	1187	5	PCT-US94-04496-28	Sequence 28, Appli

## ALIGNMENTS

RESULT 1  
US-09-103-429A-4  
; Sequence 4, Application US/09103429A  
; Patent No. 6187558  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSES: Brown, Pimisi & Michaels, P.C.  
; STREET: 118 No. 6187558th Tlloga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-3628  
; TELEFAX: (607) 256-2000  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 805 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic membrane  
; US-09-103-429A-4  
Query Match 10.4%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 8.9e-26;



```
DB 701 KKEEA-----KKEKKAAPPEEETPAKLGVEAKPKKEAEDAKA-----KEPSKP 744
OY 570 TPKEPAPTTPKGTAPTTLKEPAPTTPKKPAPKELAPTTKGPSTSTSDKPAPTTPKETAP 629
Db 745 SEKE-----KPK-----KEVPAPAEKKDOTE-----EKTESKKREKPKMEAK 784
OY 630 TPKEPAPTTPKKPAPTTPPETPPTSEVSTPTTKKEPTTIHSPDE 676
Db 785 AKED-----KGLPQEPSPKPTKEAKKSSSTDOKDSQSEKAPED 824
```

Search completed: April 26, 2002, 16:19:03  
Job time: 328 sec

RESULT 15  
 NF-H RAT STANDARD: PRT: 831 AA.  
 ID NF-H RAT  
 AC P16884; 063368;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).  
 GN NF-H OR NFH.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCB1;taxid=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain:  
 RX MEDLINE-69065087; PubMed-3143606;  
 RA Brien K.C., Robinson P.A., Wilson D., Anderson B.H.;  
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).  
 RT Identification of putative phosphorylation sites.";  
 RL FEBS Lett. 241:213-218(1988).  
 RN (2)  
 RP SEQUENCE OF 37-831 FROM N.A.  
 RX MEDLINE-88309090; PubMed-2457365;  
 RA Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,  
 RA Jolles P.;  
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and  
 RT in situ detection.";  
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).  
 RN (3)  
 RP SEQUENCE OF 1-88 AND 243-313 FROM N.A.  
 RX MEDLINE-87080760; PubMed-2878828;  
 RA Robinson P.A., Wilson D., Anderson B.H.;  
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
 RT (NF-H).";  
 RL FEBS Lett. 209:203-205(1986).  
 RN (4)  
 RP SEQUENCE OF 318-831 FROM N.A.  
 RX MEDLINE-89184647; PubMed-2928342;  
 RA Lieberburg I., Spliner N., Snyder S., Anderson J., Goldhaber D.,  
 RA Smulowitz M., Carroll Z., Emanuel B.S., Breiner J., Rubin L.;  
 RT "Cloning of a cDNA encoding the rat high molecular weight  
 RT neurofilament peptide (NF-H): developmental and tissue expression in  
 RT the rat, and mapping of its human homologue to chromosomes 1 and  
 RT 22.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: I, M,  
 CC NF-H HAS AN IMPORTANT FUNCTION IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.  
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFH IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783  
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).

CC EMBL: M37227; AAA41693.1; ALT\_FRAME.  
 DR EMBL: X13804; CAA32038.1; ALT\_FRAME.  
 DR EMBL: M21964; AAA41695.1; -  
 DR EMBL: J04517; AAA41692.1; -  
 DR PIR: A30796; A30796.  
 DR PIR: A25649; A25649.  
 DR PIR: B25649; B25649.  
 DR PIR: S02003; S02003.  
 DR InterPro: IPR001664; IF.  
 DR Pfam: PF00038; filament; 1.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.  
 FT NON REP  
 FT DOMAIN  
 FT 1 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.  
 FT CONFLICT 164 164 L -> I (IN REF. 2).  
 FT CONFLICT 185 185 I -> S (IN REF. 2).  
 FT CONFLICT 193 193 L -> T (IN REF. 2).  
 FT CONFLICT 199 199 M -> T (IN REF. 2).  
 FT CONFLICT 346 346 K -> N (IN REF. 2).  
 FT CONFLICT 373 373 A -> V (IN REF. 1).  
 FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).  
 FT CONFLICT 485 485 P -> S (IN REF. 2).  
 FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).  
 FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).  
 FT CONFLICT 727 727 A -> V (IN REF. 4).  
 FT CONFLICT 757 759 AAP -> GST (IN REF. 4).  
 FT CONFLICT 769 769 T -> L (IN REF. 2).  
 FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).  
 SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

Query Match 9.9%; Score 497; DB 1; Length 831;  
 Best Local Similarity 27.6%; Pred. No. 2.3e-13;  
 Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;  
 37 PDFTSTQHNKVTSPKITTAKPINRPSLDPNSDTSKETSITVYKETTVEKTTTNNKQ 96  
 197 PSMST--HIKKESEKIKVE-----KSEKETIVVEQVEIQTVEVTEED 242  
 97 TSDGKKTTSKETSQSIETSKDLAPTSKVLAKPPKAKETTKGALTTPREP---T 152  
 243 KEAQGEEREAEGEAAATTS-----PRAEASP---EKETKSPVKEAKSPAEKS 293  
 153 PTPKKEPA-STPKKEPTPTIKSAPTPKPEA-----PTTKSAPTPKPEA-----PTTK 203  
 294 PAAKSPAEKSPAEKSPAVAKSPAEVKSAPAEVKSAPAEKS--PAEKSPAEVKSAPATVK 352  
 204 EP-----APTPKPEA-----PTTKPEA-----PTTKSAPTPKPEA-----PTPKKPA-- 245  
 353 SPGEAKSPAEKSPAEVKSAPAEKSPAEKSPASVKS--PGEAKSPAEKSPAEVKSAPATV 411  
 246 --PTTKPEA-----PTTKPEP-----PTTKPEAPTPKPEA-----PTTKPEA-----PTA 287  
 412 KSPVEKSPAEKSPVTVSPAEKSPAEVKSAPAEVKSAPAEVKSAPAEKSPAEKSPV 470  
 288 PKKPAPTPKPEAPTPKPEAPTPKPEAPTPKPEA-----PTTKS-----APTPKPE 338  
 471 AKSPADA--KSPGAKPPEAKSPAEKSPAEKSPAEKSPAEKSPAEVKSAPAEKSP 528  
 339 APPTTSS-----APTPKPEPTTKPEA-----PTTKPEAPTPKPAAPTPKPEAPPT 389  
 529 VEGAKSLAEKSPAEKSPAEVKEIIPPAEVKSPEKSPMKREKSPKAKTLDVKSPE 588  
 390 PKPAPTPKPAAPTPKPEAPTPKPEAPTPKPEAPTPKPEAPTPKPEAPTPKPEAPTP 449  
 589 AKPPEAKKRRADIRSPQOVSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEV 644  
 450 APTPEEPPTPEEPAPTPKRAAPNTPKPEAPTPKPEAPTPKPEAPTPKPEAPTP 509  
 645 ---KAKEPKVEEETPATPKTEVKESKDEAPKEAQKP--KAKEKEPTLEKPKSPGKA 700  
 510 PKGTAPPTLKEKAPPTPKPAKELAPTTTKEPTSTSDKPAAPTPKGAAPTPKPEAPT 569



QY	92	TTNKGOTSDCKEKTTSKAKENOSIEKTSKANDLAPTSKVLAKPTPKAEITTKGRLATTPKEP	151
Db	3	TTAAATAAAGCKDTTAAABESAAAEKTA--AGVSAAPT--AAVAAGGEDATP----	51
QY	152	TTTTPEKPASTPKEPTPTTIKSAPTPPEKAPPTTKSAPPTPKKPAPTTKKEPAPTTPK	211
Db	52	-----AAATAAAEETTAAGEAPPTTTTAPRTTAAAGAPTTAAATPATAAGAPTTAT	103
QY	212	EPAPTTTKEAPAPTTTSS-APTPKKEPAPPTPKKPAPTTKKEPAPPTKEPTTPKKEPA-	269
Db	104	GKAPATAAAPVPTTAAASKAPPTTAAATSTAAAAAPTTAAASAASKERSSTSSSEEBCH	163
QY	270	--PTTKEAPPT--TPKEBAPAPKPAPTTPK-----EPAPTTKEKAPPTTKE	314
Db	164	VKRSKEMGSKGKITKKQ-----CKKNCCFDPKHGIGCHCFHNRPKSHSEEHNTTTPK-	218
QY	315	PSPTTKEPAPPTTYSAPPTTKEPAPPTTKSAPPTPKESPSTTKKEBAPPTKEPAPPTP	374
Db	219	-----APPTIOATATPPTT--TPPTT-----TTTTKATPTT-----	245
QY	375	KKPAPPTPKEPAPPTPKKEBAPPTTKKAPATPKKEBAPPTKEAPPTPKKLTPTEPKLA	434
Db	246	-----TTTTKATPTT-----TTTTKATPTT-----TPPT-----	270
QY	435	PTTPEKAPPTPELAPPTPEEPTPTTPEEAPPTPKAAAPNTPKEPAPTTKEPAPTPP	494
Db	271	-----TTTTTKATPTTTTTTTTTP-----	289
QY	495	KEBAPPTKEAPPTPKGTAPTTTKEKAPPTPKKPAKELAPTTTKEPTSTSDKAPPT	554
Db	290	-----TTTTKATTTTTTSSGECKMPSK-----REDGSGITESOCR	328
QY	555	PKG-----TAPPT-----PKEPAPPTKEPAPPTPKGTAPTTTKEKAPPTPKKPAKELA	604
Db	329	TGCGCFDSSIPDTKWCFTLSQVADCKVESQNRDCFRIT----ADOCROKNCDFSS	384
QY	605	PTTTGGPTSTSDKAPPTKEAPPTPKKEBAPPTPKKPAPTPEPTPPTTSEVSTPTT	664
Db	385	ISGTTMCVSYSSQVAA--TKTTTTPTTTTTTTTTTKATPTT-----TTTTPTT	436
QY	665	KEPTTIHKSPESTPELSAEPPTPALNSPKPEGVPTTKPATPKEMTTKDKTTEED	724
Db	437	TTTTT-----TTTA-----TTTTPTTTTTTTTT--KAT----	463
QY	725	LATPETTTAAPKMTKETATTTKTESKITATTTGVTSTTTODTTPFKITTLTKTTIAP	784
Db	464	-TTTTPTTTTTTP-----TTTTTKAT-----TTTTPTTTTTTTTTKATTTTTTTTT	510
QY	785	KVTETTKTITTEILMNKPEELAKKODATMSKATTPKPKQPTKAPKPKPTSTKPP-----T	840
Db	511	TTTTTKATTTTTTSECKME-----PSKADCGYGITESOCRSGCCFDSISIPDTKWCFS	566
QY	841	MPVAKPKPTPTPR	854
Db	567	LPOVADCKVAPSSR	580
RESULT	13		
REP1_HUMAN			
ID	REP1_HUMAN	STANDARD;	PRT; 1970 AA.
AC	P24928;		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DE	20-AUG-2001 (Rel. 40, Last annotation update)		
DR	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (REP1).		
GN	POLR2A.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92178992; PubMed=1542581;		

```

RA Maltzerich M., Acker J., Vicaire S., Vigneron M., Kedingler C.;
RT "Complete sequence of the human RNA polymerase II largest subunit.";
RL Nucleic Acids Res. 20:910-910(1992).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=95347616; PubMed=7622068;
RA Mita K., Tsuji H., Morimoto M., Takahashi E., Nenoi M.,
RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;
RL "The human gene encoding the largest subunit of RNA polymerase II.";
Gene 153:285-286(1995).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- PTM: THE TANDDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC -!- THE PHOSPHORYLATION ACTIVATES POL2.
CC -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE mRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND tRNA GENES.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63564; CAA45125.1; -.
DR EMBL; X74874; CAA52862.1; -.
DR EMBL; X74872; CAA52862.1; JOINED.
DR EMBL; X74873; CAA52862.1; JOINED.
DR EMBL; X74871; CAA52862.1; JOINED.
DR EMBL; X74870; CAA52862.1; JOINED.
DR PIR; S21054; S21054.
DR MIM; 180660; -.
DR InterPro: IPR000684; RNA_POLIII_repeat.
DR InterPro: IPR000722; RNA_POL_A_1.
DR InterPro: IPR002879; RNA_POL_A2.
DR pfam: PF00623; RNA_POL_A; 1.
DR pfam: PF01854; RNA_POL_A2; 1.
DR PROSITE; PS00115; RNA_POL_II_REPEAT; 43.
KW Transiferase: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.
FT ZN FING 71
FT DOMAIN 150 1958
FT FT 1067 1067 W->L (IN REF. 2).
FT CONFLICT 1067 1067 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
FT CONFLICT 1449 1449 D->Y (IN REF. 2).
SQ SEQUENCE 1970 AA; 217205 MW; 6876FC35692A657E CRC64;

Query Match 10.0%; Score 503.5; DB 1; Length 1970;
Best Local Similarity 33.38; Pred. No. 2.5e-13;
Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

OY 151 PPTPTKPEPASTPKAE-PTPTTKISAPPTTKPEAPTTTKSAPTTKPE-----PAPTTKKE 204
DB 1507 PSPMGISPPAMTPMNGQATPAVGAMSPVSGGMPGGAAGSPSASASDAGSPSPVSPAWS 1566
OY 205 PAPTTPKPEAPTTTKPEAPTTTKSAPTTKPEAPPTTKKAPPTTKPKDP--APTTPKEPTP 262
DB 1567 PTPGSPGSPGSSPYIPSPGAMSPSPSPSPA-YEPRSPGGYTPQSPSPSPSPSPV-SP 1624
OY 263 TTPKPEAPTTTKPEAPTTTKPEAPTPAPKPKAPPTTKPEAPTTTKPEAPTTTKKESPTTPKE 322
DB 1625 TSPSPV-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-----SPTSPSY 1671
OY 323 PAPTTPKSAPTTKPEAPTTTKSAAPTTPKESPTTPKEAPTTTKPEAPTTTKKAPPTTKPAPTTP 382

```

Db	470	PTPKRPSPPPTPKPKITTPPTPKRRKPSPTPKOKRSPPTPKKSSSPPTSKSKKTP	529
Oy	508	TT--PKGTATPTTLKEPAPTPKKRPAPKEAPLTTEPTSTSDKPAPT--TKGTAPTPPK	564
Db	530	PTPKPKITTPPTPK-----PKSPSPSYKPTKTPPTPKPKIRPPTPKPKASVPTPK	583
Oy	565	EPAPTTKEKPAPTPKGTATPTTLKEPAPTP-----KKPAPKEAPLTTPGPTSTSDKPA	620
Db	584	-----PKITTPPTPK--PKSPITPTPKOKRSPPTPKKSSSPTPAKSKKTI	628
Oy	621	PTT--PKETAPTTPEKPAPTTPK--KCAPTTPET-----PPTTSEVSTPTTKEPT	669
Db	629	PPTPKPKITTPPTPK--PKSPSPSYRKPITTPPTPKKSKSPQAKSKSPSPSYOKPT	687
Oy	670	TKSPDESTPELSAPPPKALENSKRGVPTTKTPATPKEMTTPAKDPT-----TERPL	725
Db	688	Y---PPSPYKPKITTPPTPK-----PKISPTPK-----TKSPYASKKTSVPTPKPK	735
Oy	726	RTPEPTTAAAKMKREKATPTTEKTESKITAATTTQVSTTTQDTPPKITTLTKTLAPK	785
Db	736	ISPTPKAKK-----SVPTPKPKPSYASYSKPKIRPPTPKPKPSYASAK-----PK	785
Oy	786	VTTTKTITTTTEIMNKPPEETAKPKDRATNSATTPKRO---KPKAPAKPK--TSKKPK--T	840
Db	786	I-----KPPYTKPKP-----STASSYKPKIRPPTPKPKPSYASYSKPKIT	827
Oy	841	MPVRKPKTT--PTPKMTSTMPENLPTSTIAEAMLOTTRPNO	882
Db	828	YPTPKPKISYPTPKPKITYPPTPKPKISYPTPKPKISYPSQ	871
RESULT	12		
ID	MUCL_XENLA	STANDARD;	PRT; 662 AA.
AC	005049;		
DT	01-OCT-1994 (Rel. 30, Created)		
DT	01-OCT-1994 (Rel. 30, Last sequence update)		
DT	01-OCT-1994 (Rel. 30, Last annotation update)		
DE	INTEGUMENTARY MUCIN C.1 (FIM-C.1) (FRAGMENT).		
OS	Xenopus laevis (African clawed frog).		
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;		
OC	Xenopodidae; Xenopus.		
OX	NCBI_Taxid=8355;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
TX	TISSUE=Skin;		
RX	MEDLINE=93077556; PubMed=1447205;		
RA	Hausen F., Hofmann W.;		
RT	C.P.-domains as shuffled cysteine-rich modules in integumentary mucin		
RT	C.1 (FIM-C.1) from Xenopus laevis. Polylispersity and genetic		
RT	polymorphism.;		
RL	J. Biol. Chem. 267:24620-24624(1992)		
CC	-1- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL		
CC	INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.		
CC	-1- SUBCELLULAR LOCATION: SECRETED.		
CC	-1- ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN		
CC	MAY BE PRODUCED BY ALTERNATIVE SPLICING.		
CC	-1- TISSUE SPECIFICITY: SKIN.		
CC	-1- PTM: EXTENSIVELY O-GLYCOSYLATED.		
CC	-1- SIMILARITY: CONTAINS 6 P-TYPE (TIRFOLL) DOMAINS.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
DR	EMBL; L02115; AAA74725.1; -		
DR	PIR; A45155; A45155.		
DR	HSSP; P04002; IWFA.		



```

RL      J. Mol. Evol. 43:348-356(1996).
CC      -I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC      PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GUE. THE MUSSEL'S
CC      ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC      FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC      -I- SUBCELLULAR LOCATION: SECRETED.
CC      -I- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC      -I- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC      -I- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND
CC      ALSO TYROSINE (THUS PRODUCING DOPA = 3,4-DIHYDROXYPHENYLAAMINE).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC      or send an email to license@isb-sb.ch).
CC      -----
DR      EMBL, D63777; BAA09850.1; -
DR      InterPro: IPR002964; Adhesive_pla.
DR      InterPro: IPR002965; P_rich_extensn.
DR      PRINTS: PR01216; ADHESIVEI.
DR      PRINTS: PR01217; PRICHEXTENSN.
DR      Signal: Repeat; Hydroxylation.
FT      SIGNAL          1      20
FT      CHAIN           21      872
FT      DOMAIN          21      41
FT      DOMAIN          124      872
FT      DOMAIN          184      192
FT      DOMAIN          213      221
FT      FT              SEQUENCE      872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;
SQ      SEQUENCE        872 AA; 101677 MW; 98CC70D7C75FF3C4 CRC64;

Query Match          10.3%; Score 518.5; DB 1; Length 872;
Best Local Similarity 29.2%; Pred. 3.5e-14;
Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

QY      19 VVDKSGSLDNGDEKVFNTPDIST--TOH-----NKVSTSPKITTAPRINRSLPNS 69
DB      29 VYGSVYSASAGAK-TLPGSHPGSKHPVYKPMNKIPT--PYI--SKKSYAPAPKPGY 84
QY      70 DTSKETSLTVKKEITVETKETTNTKQSTQGEKETSAAKENGOSTEKISANDLAFTSVL 129
DB      85 YPTKRYQPTGYSKNTYPIYPIAKKLSSYAKIKTTIYAIYAKTSYPSYK----HKIT 139
QY      130 AKPIPKAETTTKGALITTPKEPT--PTTPKEBASTTPKEPRTTIKSAPITPREPATTTK 188
DB      140 YPPIYK-----PKITYP--PTYKOKSPSYSPKPKTTPPYK-----PKITYPPIYK 185
QY      189 SAPT--TPKEPAPT--TTKEPAPTTPKEBAPTTPKEBATTIKSAPITTPKEBAPT--PKK 243
DB      186 RKPSVTPYKKAITYPKIYPKKITYP-----PIYKKRPST-----PYKPTTYPPYKPKI 236
QY      244 PAPITTPKEBAPT--PKPEPTTT--PKBAPTTPKEBAPTTPKEBAPT--APKKAPAPTTP 296
DB      237 SYPIIYPKKASYSSYSKSKTYRPTYPKIKISYPTYPKPKPSY--PTYKPKVITYPTTY 292
QY      297 KEPAPTTPKEBAPTTPKEBPTTPTPKEBAPTTPKAPT-----TTKEBAPTITTKSAPITTPK 351
DB      293 K-PKRSITPPIYKKKITIYPTT- YKKRPSYPTIYKOKPSIPIYKSSSYPSYKSKKTYPP 350
QY      352 EPSTTTTKEPA--PTTPKEBAPTTPKKP-APT-TPK-----EBAPTTPKEBAPTTPKKP 401
DB      351 TYKKKITIYPTIYKPKPSYPSYKPKKISYPTYPKIKIYPTIYKPKPSYPSYKSKKTYPP 410
QY      402 APAPKPEAPATTPKEPAP-----TTPKKLPTTPPEKLAAPITTPKEPAPTTEELAPITPEE 456
DB      411 T-YKPKISYPTIYKPKRSYSSYSKSKTYPPIYKPKIKISYPTIYKPKRSYPTIYKPKKITYP 469
QY      457 PT--PTTPPEBAPTTPKAADNTPK-EBAPTTPKEBAPTTP-----KEBAPT--PKETAP 507

```

```

DE CALPHOTIN.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydraidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RL "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RL "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC OF CA+2 PER MOL OF PROTEIN.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC COMPOUND EYES AND OCELLI.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC DEVELOPMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial/
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/sib.ch).
CC or send an email to license@sib-sib.ch.
CC -----
CC EMBL: L02111; AAA28405.1; -.
CC EMBL: L05080; AAA28420.1; -.
CC PIR: A47282; A47282.
CC Flybase: FBgn0010218; Cpn.
KW Calcium-binding.
FT CONFILCT 36 36 A -> AVAPAVVA (IN REF. 2).
FT CONFILCT 43 43 I -> T (IN REF. 2).
FT CONFILCT 64 64 I -> V (IN REF. 2).
FT CONFILCT 76 76 T -> A (IN REF. 2).
FT CONFILCT 100 100 P -> PP (IN REF. 2).
FT CONFILCT 126 127 VO -> AP (IN REF. 2).
FT CONFILCT 154 154 I -> V (IN REF. 2).
FT CONFILCT 160 160 S -> T (IN REF. 2).
FT CONFILCT 534 534 A -> E (IN REF. 2).
FT CONFILCT 699 699 I -> T (IN REF. 2).
FT CONFILCT 703 703 V -> L (IN REF. 2).
FT CONFILCT 721 721 D -> E (IN REF. 2).
FT CONFILCT 865 AA: 84781 MW: 2110417E0B0E7CFE CRC64;
SQ SEQUENCE

Query Match 10.6%; Score 530; DB 1; Length 865;
Best Local Similarity 26.2%; Pred. No. 1.2e-14;
Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 411

OY 123 APTSKVLAKD-TPK---AETTTKGPAITTPKKEPTPTTPKPEASTTTPKKEPTTIKSAPTT 178
Db 8 SPSPASVAEAVETPSAVAAPOQVSPVAAPV-----APAPAIIVTVPVAPPPTLASVOPAT 61
OY 179 PKEPAPTTTSAPTTP-KEPAPTTTKEPAPTTTKEPAPTTTKEPAPT-----TTKSAPTTP 233
Db 62 VTIPAPAPIAAASVTIPVASVAPPVAAFP-----PAASVSTIPVAAQIIPVASAPVAP 116
OY 234 KEPAPTPPKKPAP-----TTPKEPAPTTTPKKEPT-----TTPKEPAPTTKEPAPT 279

```

DR	Pfam: PF00038; filament; 1.	
DR	PROSITE; PS00226; IF; 1.	
KW	Intermediate filament; Coiled coil; Neurone; Phosphorylation;	
KW	Repeat.	
FT	DOMAIN	1
FT	DOMAIN	97
FT	DOMAIN	98
FT	DOMAIN	408
FT	DOMAIN	409
FT	DOMAIN	1087
FT	DOMAIN	436
FT	DOMAIN	517
FT	DOMAIN	519
FT	DOMAIN	886
FT	DOMAIN	887
FT	DOMAIN	1087
FT	DOMAIN	98
FT	DOMAIN	129
FT	DOMAIN	141
FT	DOMAIN	130
FT	DOMAIN	142
FT	DOMAIN	239
FT	DOMAIN	240
FT	DOMAIN	261
FT	DOMAIN	262
FT	DOMAIN	283
FT	DOMAIN	287
FT	DOMAIN	288
FT	DOMAIN	408
FT	CONFLICT	133
FT	CONFLICT	133
FT	CONFLICT	199
FT	CONFLICT	281
FT	CONFLICT	281
FT	CONFLICT	492
FT	CONFLICT	492
FT	CONFLICT	551
FT	CONFLICT	551
FT	CONFLICT	689
FT	CONFLICT	712
FT	CONFLICT	714
FT	CONFLICT	714
FT	CONFLICT	814
FT	CONFLICT	814
FT	CONFLICT	843
FT	CONFLICT	843
SO	SEQUENCE	
	1087 AA; 116612 MW; 57BAC76A38ED1CB9 CRC64;	

Query Match	10.6%	Score 533	DB 1	Length 1087
Best Local Similarity	27.5%	Predg NO. 1.1e-14		
Matches	216	Conservative 81	Mismatches 330	Indels 158
				Gaps 38

```

0Y 21 DEAGSSLONGDGRV-----PDTSTOHNKSTSPKITTAKPILNPROSLPNDSOTKET 75
Db 404 EECRIGFGSPSLTEGLKIPISIT--HIKXSEEMIKVTE-----KSEKET 449
0Y 76 SLTVNKETIVERKETTTTKOTSTDGKEKTTSAKETSIOSEKTSAKDLAPTSVLKLPMPX 135
Db 450 VIVEGOTEEIRNTEGVTEEDKEKAGOGEEEAEBEGEKEEELLAATSPPAEANSPEKE 509
0Y 136 AETTTKGPAITTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTITTKASAPTPP 195
Db 510 TTSRVKEEAKSGEAKSGEAKSPA-----EAKSPGEAKS--PGEAKSPGEAKSPAPKPSA 564
0Y 196 BP-APTTPKEPAPTTPKEPAPTITTEP-----APTTKSAPTTPKCPA-----PTTKKPA- 245
Db 565 EKSSPAEEKSPA--BKSPA--TVKSPAEANSPPSAKS--PAEAKSPAARKSPAEAKSPA 619
0Y 246 ---PTTPKEPA-----PTTPKEP-----PTTPKEPAPTTPKEPAPTTPKEPA-----PTAPKK 290
Db 620 AKSPAEAKSPAEAKSPAIVKSGEAKSPSEAKSPA---EAKSPAIAKSPAEEKSPA 676
0Y 291 P-----APTTPKEPAPTTPKEPAPTITTEPPTTPKEPA-----PTTKSAPTTPKEPAPT 342
Db 677 PGEAKSPAEPKSPA--KSPA---EYKSPAEAKSPAIVKSPGEAKS--PAAVKSPAIAK 729
0Y 343 TKSAPTTPKE--PSPTTPKEPA-----PTTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPT 397
Db 730 SPAVKSNGEAKSPGEAKSPAIAKSPAIAKSPAEVKSPEKKTTPVKEGAKSPAIAKSPBEK 789
0Y 398 TKKPAAPTAPKBPAPTTPKETAPTTPKLTPTTPKELAAPTTPKEPAPTTPPELAAPTTPPEP 457
Db 790 AKSPVKEDIKPPAEAKSPEKA--KSPVKEGAKBPPEKAKLDVKSBEAQTPVOEBAETVPDI 848
0Y 458 TPTTPPE-PAPTTPKAAAAPTTPKEPAPT---KPEPAPTTP-----KEPAPTTPKET 505
Db 849 RP--PEOYKSPAIAKSPA--KEEAKTSEKVAAPKEEVKSPVKEEVAKBPAPKVEBEK 904
0Y 506 APTTPKGAAPTTPKEPAPTTPKPAKELAPTTPKEPTSTSD--KAAPTTPKGTAPTTP 563
Db 905 TLTPPKPAEAKSKDDEAPKEAPKPAVEEKTTPTEKPPDOSTAIEAKKKEBAGEKKKVAASEE 964

```

[illegible]

ID	TCNA.TRYCR	STANDARD:	PRT:	1162 AA.
DT	P23253;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).			
CC	TCNA.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma,			
OX	NCBI_Taxid=5693;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SIIVIO X-10/4;			
RX	MEDLINE=91277609; PubMed=1711561;			
RA	Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,			
RA	Prioli R.P.;			
RT	"The Trypanosoma cruzi neuraminidase contains sequences similar to			
RT	bacterial neuraminidases, WTD repeats of the low density lipoprotein			
RT	receptor, and type III modules of fibronectin."			
RL	J. Exp. Med. 174:1179-191(1991).			
RN	[2]			
RP	SUBCELLULAR LOCATION.			
RX	MEDLINE=91376547; PubMed=1896773;			
RA	Prioli R.P., Mejia J.S., Ajl T., Aikawa M., Pereira M.E.A.;			
RT	"Trypanosoma cruzi: localization of neuraminidase on the surface of			
RT	trypanositigotes."			
RL	Trop. Med. Parasitol. 42:146-150(1991).			
CC	- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN			
CC	PARASITE INVASION OF CELLS.			
CC	- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC			
CC	LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETINEURAMINYL			
CC	RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACETLATED			
CC	NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,			
CC	GLYCOLIPIDS OR COLOMIGIN ACID.			
CC	- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR			
CC	(POSSIBLE).			
CC	- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPANOSITIGOTES, MINIMUM			
CC	IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.			
CC	- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT			
CC	DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA			
CC	PROTEIN.			
CC	- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.			
CC	- SIMILARITY: CONTAINS 3 BNR REPEATS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
OR	EMBL; M61732; AAA30255.1; -			
OR	PIR; JH0557; JH0557..			

-1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTADEPTIDE  
SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
GLYCOSYLATED.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X13885; CAA32090.1; -  
DR PIR: S06733; S06733.  
KW Repeat: Cell wall; Glycoprotein; Signal; Structural protein;  
KW Hydroxylation.  
FT SIGNAL 1 ?  
FT CHAIN 70 620 EXTENSIN.  
FT REPEAT 73 H-A-P-P.  
FT REPEAT 148 151 H-A-P-P.  
FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.  
FT REPEAT 229 235  
FT REPEAT 236 242  
FT REPEAT 205 620  
FT DOMAIN 499 600  
FT DOMAIN 620 600  
SQ SEQUENCE 620 AA; 65406 MW; 64IDD2278AB28524 CRC64;

Query Match 11.0%; Score 551; DB 1; Length 620;  
Best Local Similarity 27.2%; Pred. No. 1,5e-15;  
Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY 136 AETTK-GPALTP--KEPTT-----PKPASTPKK-----PPTTKASAPTTKEPAP 184  
DB 24 AEATTOYGYLPRVTSQPPSSIGLSPPSAPTTTPSRGHNVPSP---RHAPRHAYPP 80  
QY 185 TTKTSAPTTKEP-----APTTKEKAPAT---TPKEAPATTTKEAPATTTKSAPTTKE 235  
DB 81 SHGLPRPSVGGPRHGRHLPRSGFNRPSPVYISPSHNPESYGAPRPSHGCHLPSHCQR 140  
QY 236 PAPTTPKPPAT---TPKEAPATTTKEPAPATTTKKEPAPATTTKKEPAPATTTK 290  
DB 141 P-----PSPSHGNAPRSGGTHPRGONHRRSPSRHGNHPRPTTYAQRPPTIYS 193  
QY 291 PAPTTPKPPATTTKEPAPATTTKERSPTTK---EPATTTKSAPTTTKEPAPATTTK 345  
DB 194 PSPGV--QRPPT--YSPRPPTHVQTPSPSRGHQOPRTHHAPRTHHAPRTHHOSPL 249  
QY 346 ---APTTPKESPPTTKEPAPATTTKKEPAPATTTKPPATTTKEPAPATTTK 399  
DB 250 RHLPPSPRKOQPTYSPPRAYAQSOPRSTYSPPPTYSPPPTYSPPRAYSPSP 309  
QY 400 KPAAT---APKEAPATTTKETAPTTPKKL-TPPT-----PEKAPTTKEPAPTTPPEL 449  
DB 310 PPTLPPTSPPPAYSPPTYSPPPTTYLPLPSSSIYSPRPVYSPPPPSYSPPTTYL 369  
QY 450 APPTPEPTPTTKEPAPATTTKAAAP---NTPKEAPATTTKEPAPATTTKKEPAPATTTKETA 506  
DB 370 PPPSPSPSPSPSPPTTYSQSPPPRAYSPPLAPPTYSPPPT--YSPPPPTIYAQP 427  
QY 507 PTTKGTAPTTLKEPAPATTTKPPAKELAPTTTKEPTTSDKRAPATTTKGTAPTTPKEP 566  
DB 428 P-----LPPTYSPPRAYSPPTTYSPPPTYSPPRAYAQPAPPPPTYSPPRAYSP 481  
QY 567 APPTTPKEAPATTTKGTAPTTLKEPAPATTTKPPAKELAPTTTGTSTSDKRAPATTTKE 626  
DB 482 PPPSPYSPPPYQVPLRPTSPSPRRRIHLRPPRHOP---RPPTTYGQPPSPPTSP 538  
QY 627 TAPTTPKPPAT---TPKAPATTTPTPTPTSEVSTPTTKEPTTIHSDSTSELSAE 684  
DB 539 PPPROIHSPPRHMOQRIPTTYGQPPSPPTFSAPPPROIHSPPRHMOQRPRTTYGQ 598  
QY 685 PTPKALENSPKRPGVPTTKTPAATKP 710

DB 599 PSP-----PTTSPSPSP 612

RESULT 8  
ID NPH\_MOUSE STANDARD; PRT; 1087 AA.  
AC P19246; 061959;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).  
GN NPH OR NPH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121513; PubMed=3220257;  
RA Jullien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,  
RA Mushynski W.;  
RT "sequence and structure of the mouse gene coding for the largest  
RT neurofilament subunit.";  
RL Gene 68:307-314(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89089138; PubMed=3145094;  
RA Sheldman P.S., Garden M.J., Lees J.F., Lazzarini R.A.;  
RT "The structure of the largest murine neurofilament protein (NF-H) as  
RT revealed by cDNA and genomic sequences.";  
RL Brain Res. 464:217-231(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;  
RA Carden M.J.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
CC OBSERVED BY THE TWO SMALLER NF PROTEINS.  
CC -1- PHOSPHORYLATION: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P. NPH IS  
CC THOUGHT THAT PHOSPHORYLATION OF NPH RESULTS IN THE FORMATION OF  
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
CC OF AXONAL CALIBER.  
CC -1- PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534  
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M24496; AAA39813.1; -  
DR EMBL: M23499; AAA39813.1; JOINED.  
DR EMBL: M24494; AAA39813.1; JOINED.  
DR EMBL: M24495; AAA39813.1; JOINED.  
DR EMBL: M35131; AAA39809.1; ALT\_FRAME.  
DR EMBL: Z31012; CAA83229.1; -  
DR PIR: JTO368; QFM5H.  
DR PIR: A43778; A43778.  
DR MGD: MGI:97309; NfH.  
DR InterPro: IPR001664; IF.

Query Match	11.1%;	Score 555.5;	DB 1;	Length 2700;
Best Local Similarity	32.6%;	Pred. No. 3e-15;		

	Matches	199:	Conservative	64:	Mismatches	240:	Indels	107:	Gaps	32:
Qy	58	PINPRSLPNSTOSKETSLTVNKEFTVETKEETTTNNKOTSDGKEKTTSAKETOSIERT								
Db	445	PVVLDELPEVSEVSS-----TGSEPTGLTENPTSTK-----KPTVSIERP								
Qy	118	SAKDLAPTSVLAKPTPKAETTTGKALTPPKPPPTPKPKAPPTPKPT-----PTTI								
Db	488	SVYTEKPT-----VPREKPTIPTEKTTSTEK---PTISEKPNMPSEKPTIPSEKPTIL								
Qy	173	KSAPTPKKBPAPTTTKSAPPTPKKBPAPTTTKKBPAPTTTKKBPAPTTTKKBPAPTTTKSAPPT								
Db	540	TEKPTIPSE-KPTIPSEKPTISTEKPTVPEE--PTTPEEFTTVMEEVPTKPSIPT								
Qy	232	TPKEAPPTPKKBPAPTTPKKBPAPTTPKKBPAPTTPKKBPAPTTPKKBPAPTTPKKBPAPTTPK								
Db	597	--EKSEIPEK---PTSMEEITISTEKPTICEPEKPTIPEK---PTIPEKSTISPEK-								
Qy	292	APPTPKKBPAPTTPKKBPAPTTKPSPTPKKBPAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTPK								
Db	648	-PTTPTPE-KPTIPEKPTISTEKPTIPEK-PTISEKTLIPEKTLIPT--EKPTIPT								
Qy	352	EPSPPTTKKBPAPTTPKKBPAPTTPKKBPAPTTTKKBPAPTTPKKBPAPTTTKKBPAPTTTKKBPAP								
Db	702	EKPTISTEE--PTTPEEFTTISTEKPSIPMEKPTIPEEFTTISVEETISTEKTIIPM--								
Qy	407	KEPAPTTKETAPT-----TPKKLPTTPEKLA-----PTTPEKBPAPTTPEEAPTTPEE								
Db	758	EKPTISTEKPTIPEKPTISTEKTLIPT-EKTLIPEKPTIPIEFTTISTEKL-TIPE								
Qy	457	PTPTPEEAPTTPKRAAPNTKEBPAPTTPKKBPAPTTPKKBPAPTTPKKBPAPTTPKKBPAPTTPKGAPT								
Db	815	KPTISEKPTISTEK-----PTIPE-KPTIPE-----ETTISTEKTLIPT								
Qy	517	TKBPAPTTPKKBPAPTELAPTTTKEPTSTY-----SDKAPTTPKGAPTTPKBPAPTT								
Db	856	--EKPTISEKTLIPEKPTISTEKPTIPEKTLIPTKPTIPEKPTIPEKPTIPEKPTIPEKPTIPEKPTI								
Qy	571	PKBPAPTTPKGAPTTLKEBPAPTTTKKBPAPTELAPTTTGGPTSTTSOKPAPTTKET--A								
Db	912	PHPSPTA-TGIALVMSPHAPSTPMTSV---ILGTTTSSSTGMSCP-PNARESCAC								
Qy	629	PTTPEKAPT	638							
Db	967	PASCSPRS	976							

RT protein.";  
 RL Biotechnol. Prog. 6:171-177(1990).  
 RN PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RP MEDLINE=8135732; PubMed=6298211;  
 RX Waite J.H.;  
 RA "Evidence for a repeating 3,4-dihydroxyphenylalanine- and  
 RT hydroxyproline-containing decapeptide in the adhesive protein of the  
 RL mussel, Mytilus edulis L.";  
 RL J. Biol. Chem. 258:2911-2915(1983).  
 CC -I- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC -I- SUBCELLULAR LOCATION: SECRETED.  
 CC -I- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC -I- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
 CC -I- PTH: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY  
 CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE  
 CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE  
 CC (DOPA) DERIVED FROM TYROSINE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X54422; CA38294.1;  
 DR InterPro: IPR002964; Adhesive\_p1a4.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PR01216; ADHESIVEI.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR Repeat: Hydroxylation.  
 FT NON\_TER 1  
 FT DOMAIN 67 870  
 FT TANDEN REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-  
 FT P-P-[ST]  
 FT SEQUENCE 875 AA; 100412 MM; 6EAB5312748CAACE CRC64;  
 SQ  
 Query Match 11.1%; Score 556.5; DB 1; Length 875;  
 Best Local Similarity 28.7%; Pred. No. 1.2e-15;  
 Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;  
 42 TQHNKYSTSPKITAKPINP-----RP--SLPNSDSKETSILTVEKETTET 87  
 1 TKEHPYKPKTSYSAHYKRPYQPLKKVYDPTKSYPTYG-SKTNVLPPLAKLSSYK 59  
 88 KETTTNKTSTDG--KEKTT---SAKETOSIEKTSKDLAPTSKYLAKEPTPAETTTG 142  
 60 PIKTYNAKNNYPPVYKPKMTYPPYKPKPSYPTYSKPKTYPKITTYPTTKAKSY-- 117  
 143 PALTTPKEPTTPKE-----PASTTPKEPTPTTIKSAPTP-----KEPATTTTKSAPTT 193  
 118 PSSYKPKKTPYKTKLTPYKPKPSYPTYSKPKPSYPTYSKPKPSYPTYSKPKPSY 177  
 194 PKEPAPTTTKEPA--PTTKEPA-----PTTKEPA--PTTTSKAPTTPK--EPAPTT 240  
 178 P-----PTYKAKPSYPTYSKPKPSYPTYSKPKPSYPTYSKPKPSYPTYSKPKPSY 233  
 241 PKKPA--PTTPKEPA--PTTPKEPT--PTTPKEP--APTTPKEAPTTTKEPA--PTAPK 289  
 234 KAPSPYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 293  
 290 KPA--PTTKEPA--PTTKEPA--PTTKEPS--PTTKEPA--PTTTSKAPTTTKEPA-- 339  
 294 KPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 353  
 340 PTTTSKAPTTKEPSPTTTPKEAPTTTKEAPTTTKEAPTTTKEPA--PTTKEPA--PTTPKEAP 395  
 354 PSYTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 409

QY 396 TTTKKPAPTAKEPAPTTPKEAPTTPKLTPPTPEKLAPTTPKEAPTTPEELAPTTPE 455  
 DB 410 PPTTKAKPPTTKAKP-----TYPST-YKAKPSY-----PSYKAKPSYPTTKAKPSY 457  
 QY 456 EPTPTPEEPAPTTPKKAAPNTPKPEAPTTTKEPAPTTKE-----PAPTPEKTAAPT-- 509  
 DB 458 KPT-----YPSYKA-----KPSYASYSKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 506  
 QY 510 PKGTAPTLKEAPPTTPKPAKELAPTTTKEPTSTSDKAPPTPKGTAPTT--PKE 565  
 DB 507 PLYTKAPTYK-PKPSYPSYKPKTYPTTK-----PKTSYPTTKAKPSY 551  
 QY 566 PAPTPKEAPTTTKEPA--APTTLKEPA--PTTPKKA--PKELAPTTTG-PTSTTS 617  
 DB 552 PATTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 611  
 QY 618 KPAPTPKEAPTTTKEPA--PTTPKKA--PTT-----DETPPTTSYSTPTTKE----- 666  
 DB 612 KAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 667  
 QY 667 PTTTKSPDESIPPELSAETPKALENSK-----EGVPTT--KTPA-----ATKPEMTT 715  
 DB 668 PPT-YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 726  
 QY 716 AKDKTERDLRTPTETTTAPKMTKETATTEKTESKITATTGVTSTTODTTPFKIT 775  
 DB 727 YKAPPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 777  
 QY 776 TLTQTTIAPKVTYTKTITTTTEIMNKPEETAKPKDRAKNSKATTPKPKPTAPK--P 832  
 DB 778 TYKSKSIYPSYKPKKTPPT--YKPKLTPPTTK-----PKPSYPSYKPKKTIY 826  
 QY 833 TSKKTKKTMPRKPKKTPPT-----RKMTSTMPLENTPSR 868  
 DB 827 STYKPKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 874  
 RESULT 6  
 ZAN\_HUMAN STANDARD; PRT: 2700 AA.  
 ID ZAN\_HUMAN  
 AC Q9Y493; O00218;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ZONADHESIN (FRAGMENT).  
 GN ZAN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE OF 1-2379 FROM N.A.  
 RX MEDLINE=99018118; PubMed=9799793;  
 RA Glockner G., Scherier S., Schatnevoy R., Boiligt A., Weber J.,  
 RA Tsui L.C., Rosenthal A.;  
 RT "Large-scale sequencing of two regions in human chromosome 7q22:  
 RT analysis of 650 kb of genomic sequence around the EPO and CMT1 loci  
 RT reveals 17 genes";  
 RL Genome Res. 8:1060-1073(1998).  
 RL [1]  
 RN [1]  
 RP SEQUENCE OF 2338-2700 FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=97271566; PubMed=9126492;  
 RA Gao Z., Harumi T., Garbers D.L.;  
 RT "Chromosome localization of the mouse zonadhesin gene and the human  
 RT zonadhesin gene (ZAN).";  
 RL Genomics 41:119-122(1997).  
 CC -I- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA  
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR  
 CC SIGNALING.  
 CC -I- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE

DR EMBL: M61170; AAB53150.1; -.  
 DR EMBL: X52229; CAA36478.1; ALT\_SEQ.  
 DR EMBL: X52228; CAA36477.1; ALT\_SEQ.  
 DR EMBL: M35093; AAB59612.1; ALT\_SEQ.  
 DR EMBL: 217324; CAA78972.1; -.  
 DR EMBL: 217325; CAA78973.1; -.  
 DR EMBL: M31823; AAA35757.1; -.  
 DR EMBL: S81781; AAD14376.1; ALT\_INIT.  
 DR EMBL: S81736; AAD14369.1; ALT\_INIT.  
 DR EMBL: M21868; AAA59874.1; ALT\_SEQ.  
 DR PIR: A35175; A35175.  
 DR PIR: B35175; B35175.  
 DR PIR: S10218; S10218.  
 DR GLYCOSULEDB: F15941; -.  
 DR MIM: 158340; -.  
 DR MIM: 113720; -.  
 DR InterPro: IPR000082; SEA.  
 DR Pfam: PF01390; SEA: 1.  
 DR SMART: SM00200; SEA: 1.  
 DR PROSITE: PS50024; SEA: 1.  
 KW Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;  
 Repeat; Alternative splicing.  
 FT SIGNAL 1 23  
 FT CHAIN 24 1255  
 FT DOMAIN 24 1162  
 FT TRANSMEM 1163 1186  
 FT DOMAIN 1187 1255  
 FT DOMAIN 81 960  
 FT DOMAIN 1034 1151  
 FT CAROHD 957 975  
 FT CAROHD 975 975  
 FT CAROHD 1029 1029  
 FT CAROHD 1055 1055  
 FT CAROHD 1133 1133  
 FT VARSPLIC 19 19  
 FT VARSPLIC 20 22  
 FT VARSPLIC 20 31  
 FT VARSPLIC 126 905  
 FT VARSPLIC 1077 1087  
 FT VARSPLIC 1088 1255  
 FT CONFLICT 2 2  
 FT CONFLICT 134 134  
 FT CONFLICT 154 154  
 FT CONFLICT 1021 1021  
 FT CONFLICT 1251 1251  
 SQ SEQUENCE 1255 AA; 122072 MM; 5E28DFC4DE7D9A82 CRC64;

Query Match 13.0%; Score 651; DB 1; Length 1255;  
 Best Local Similarity 27.1%; Pred. No. 3.4e-19;  
 Matches 290; Conservative 79; Mismatches 437; Indels 266; Gaps 52;

QY 75 TSLVNETVETETKETTNTKQSTGDKETTSKAKETOSIEKTSAND-LAPTSKVLAKPT 133  
 DB 16 TLVLV-----VSGSGHASTPGKEKETTSATCRSSPSSTSEKNAVMTSSVLSHS 65  
 QY 134 P-KAETTTKGP--ALUTPKP-----PTTKEKPASTP-----KE 166  
 DB 66 PGSSSTTGGDDVLAIPATERPASASATNGODVTSVPVTPALGSTTPRAHDVTSAPDNK 125  
 QY 167 PTPPT-----IKSAFTTKEKPAFTT-----KSAPTTKEKPAFTTKEP-----A 206  
 DB 126 PAFGSTAPRAHGVISADT--RPAFGSTAPRAHGVISADT--RPAFGSTAPRAHGVIS 181  
 QY 207 PTPKEKPAFTTKEKPAFTTTSKAPTTKEKPAFTT-----KKPAFTTKEP----- 252  
 DB 182 PDT--RPAFGSTAPRAHGVISADT--RPAFGSTAPRAHGVISADT--RPAFGSTAPRAHGV 238  
 QY 253 --APTTPKEPTTPP-----KEPAFTTKEP-----APTTPKEKPAFTTAPK----- 289  
 DB 239 TSA PDTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGV 298

QY 290 KPAFTTKEKPAFTTKEP-----EPAPTTKEKPSPTTKEKPAFTT-----TKSAPTTKEKPAFTT 342  
 DB 239 TSA PDTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGV 358  
 QY 343 TKSAPTTKEKPSPTTKEP-----APTTPKEKPAFTTKEP-----KPAFTTKEKPAFTTKEP 391  
 DB 339 T-SADT--RPAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRA 415  
 QY 392 --EPAPTTTKEKPAFTTKEKPAFTTTPKTAFTTTPKTKLTPREKLAFTTTPKPAFTTPEE 448  
 DB 416 HGVTSAPDTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRA 473  
 QY 449 LAPTTPEEPPTTPEEPAPTTPKA-----AAPTTKEKPAFTTTPK-----EPAPTTKEKPAFTT 500  
 DB 474 --PAHGVTSADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGST 531  
 QY 501 TPKEAPTTTPKTAFTTTPKTAFTTTPK-----KKPAKELAPTTTKEPTTSDKP 550  
 DB 532 AP-----PAHGVTSAPDTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPR 586  
 QY 551 AP--TPKGAFTTTPKKEPAPTTTPKPG-----TAPTTTKEKPAFTTTP----- 595  
 DB 587 APGSTAPRAHGVIS-----APDTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISAP 642  
 QY 596 -KKPAKELAPTTTKEKPTTSDKRAP--TPPK-----ETAPTTKEKPAFTTTP----- 640  
 DB 643 DTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISAP 701  
 QY 641 --KKPAFTTPEEPPTTSEVSTPTTKEPTT-----IKSPDESPTPELSAPTTKALE 691  
 DB 702 PDTTPRAFGS--TAPRAHGVTSADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGV 759  
 QY 692 NSPKPEPVPTTKPRAA-----TKPEMTTAKDK--TTERDLTPTEPT-----T 733  
 DB 760 SAPDTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGVISADTTPRAFGSTAPRAHGV 819  
 QY 734 AAKPKETETTTEKTESKITATTOVSTTQODTP--FKTTTLAKTTLAPKVTTPTK 791  
 DB 820 SAP-----DTPRAFGSTAPRAHGVTSADTTPRAFGSTAPRAHGVTSADTTPRAFGSTAPRA 875  
 QY 792 TITTT--ELINKKEETAKPRDRTAKTSKATTPKQKPTKAKKPTTSKKKTPRVAKPKT 849  
 DB 876 HGVTSAPDTPRAFGSTAPRAHGVISADTTPRAFGS--GSTAPRAHGVISADTTPRA--APGS 930  
 QY 850 TTPPKMTSTMPRLN-----TSRIEAMIQ--TTTRPQTPNSK-- 887  
 DB 931 TAPRAHGVTSAPRNPALCSTAPRVNNTVSASGSASGSTLVHNGTSARATTPRAKST 990  
 QY 888 -----LIVENPKSBDAGAGETPPHMLLRPHVFMPEVTDMOY 925  
 DB 991 PFSIPSHHSDTPPTLASHSTKTDASSTHNSVRLPSSHNSHSTPOLSTGVST 1042

RESULT 5  
 FPI\_MYTED STANDARD: PRT; 875 AA.  
 AC Q25460;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT  
 DE PROTEIN 1) (MEFPI) (FRAGMENT).  
 GN FPI.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
 OC Mytiloidae; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP  
 RA MEDLINE=91025829; PubMed=1367451;  
 RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;  
 RT "Structural and functional repetition in a marine mussel adhesive







```

OY 492 TPKKEPAPT-TPKEIAPTTPKGTAPTTIKKEPAPTTPKKAPKELAPTTKEPTSTSDKP 550
DB 1950 TTTVTTPPTGQTPTPTTTTTPPTTTPPTGQTPTPTTTTTPPTTTPPTGQTPT 2009
OY 551 AP-TTPKGTAPTTKKEPAPTTPKKEPAPTTPKGTAPTTIKKEPAPTTPKKAPKELAPTTK 609
DB 2010 TPTTTPPTTTTTPPTTTPPTGQTPT-TPTTTPPTTTPPTTTPPTGQTPTTPTTTP 2068
OY 610 GPTSTTSDDKAP-TTPKEIAPTTPKKEPAPTTPKKAPTTTETPTPTTSEVSTPTTKEPT 668
DB 2069 TPTPTPTGQTPTPTTTTTPPTTTPPTGQTPT-TPTTTPPTTTPPTTTPPTGQTPT 2127
OY 669 TIKHSPESTPELSAEPTEKALENSPEK-----GVPT-TKTPATKPEMTTAK 717
DB 2128 T-TPTTTPPTTTPPTTTPPTGQTPTTPTTTPPTTTPPTTTPPTTTPPTTTP 2184
OY 718 DKTERDLR--TTP-ETTTAPKMT--KEIATTEKTESKIATTTQVSTTTQDT 769
DB 2185 TPTPTPTGQTPTPTTTTTPPTTTPPTGQTPTTPTTTPPTTTPPTTTPPTTTP 2243
OY 770 TPKKITLKTTLAPKVTTP-KKITTELMNKKEETAKPKDRATNSKATTPKPKQKPTKA 828
DB 2244 TP--ITT--TTTTPPTTTPPTGQTPTTPTTTPPTTTPPTTTPPTTTPPTTTP 2299
OY 829 P-KKPTSTKPKKTPRVRKPKTTPPKRMSTMBELNP--TSRIAEMLQTTTR-PNQT 883
DB 2300 TPTPTPTGQTPTPTTTTTPPTTTPPTGQTPTTPTTTPPTTTPPTTTPPTTTP 2359
OY 884 PMSKLVENPKSESDAGAEGETPHMLRPVHFMPEVTP 921
DB 2360 PTTTTTTPPTTTPPTG--TQTP--TTTTTTTTTTP 2392

RESULT 2
SLIP1_CLOTM STANDARD: PRT: 1664 AA.
ID SLIP1_CLOTM STANDARD: PRT: 1664 AA.
AC Q06852;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER PROTEIN 1).
GN OLPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
OC NCBI_TaxID=1515;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=NCIB 10682;
RC MEDLINE=93209931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein Cipa and a protein possibly involved in attachment of the cellulosome to the cell surface.";
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993)
CC -1- SUBUNIT: ASSEMBLED INTO MONO LAYERED CRYSTALLINE ARRAYS.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOMOLOGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X67506; CAA47841.1;
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH. 3.
CC PROSITE: PS01072; SLH_DOMAIN. 2.

```

```

KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
FT DOMAIN 1378 1449 T-P-S-D-E-P.
FT DOMAIN 1453 1494 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1495 1565 SLH 1 (INCOMPLETE).
FT DOMAIN 1566 1635 SLH 2.
FT DOMAIN 1626 1646 SLH 3.
FT DOMAIN 1664 178194 SLH 4 (INCOMPLETE).
SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9E74B CRC64;

Query Match 18.8%; Score 942; DB 1; Length 1664;
Best Local Similarity 33.0%; Pred. No. 2,2e-30;
Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

OY 128 VLAKPRP-KAETTKGPAITPKKEPTTTPKKEPASTTPKKEPTTTIKSAPTTPKKEPAPT 186
DB 758 VLIQAPAIKAASDEPIPTDPSDEPTPS-----DEPTPS--DEPTPSDEPTPSD 804
OY 187 TKSAPTTPKKEPAPTTPKKEPAPTTPKKEPA--PTTKEPAPTTPKSAPTTPKKEPAPTTPK 242
DB 805 EPTSETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 864
OY 243 KPAITTPKKEPAPTTPKKEPTTPKKEPAPTTPKKEPAPT--TPKEPAPTAKKAPATTPKKEPAP 301
DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--DEPTPSDEPTPSDEPTPSDEPTPS 919
OY 302 TTPKEPA--PTTKEPAPT--TPKKEPAPTTPKS--APTTPKKEPAPTTPKSAPTTPKKEPAPT 357
DB 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS-----DEPTP 968
OY 358 TKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPT--TPKKEPAPTTPKKEPAPTTPKKEPAPT 416
DB 969 SDEPTPS--DEPTPS--DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD-- 1019
OY 417 TAPTPPKKLITTPPKLAPTTPKKEPAPTTPKKEPAPTTPKEEPTTP--TPEEAPAPTTPKAAAP 475
DB 1020 -----EPTPSDE--PTPSDEPTPSD--EPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1064
OY 476 NTPKEPAPTTPKKEPA--PTTKEPAPT--TPKEIAPTTPKGTAPTTIKKEPAPTTPKKAPK 532
DB 1065 TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS--DEPTPS 1122
OY 533 ELAPPTTKEPT-STTSDDKAPATTPKGTAPTTTPKKEPAPTTPKKEPAPTTPKGTAPTTIKKEPA 591
DB 1123 D-EPTPSDEPTPSSETPPEEPIPTDPSDEPTPSDEPTPS--DEPTPS--DEPTPSDEPT 1175
OY 592 PT-TPKKAPKELAPTTTTPKGTSTSDKAPAPT--TPKEIAPTTPKKEPAPT--TPKKAPAT 646
DB 1176 PSETEEP-----IPTDPSDEPTPSDEPTPSDEPTPSD--EPTPSDEPTPSSETPPEEPIPT 1229
OY 647 TPEPTPTTSEVSTPTTTPKEPTTIKHSPESTPELSAEPTEKALENSKEPCVPTTKTPA 706
DB 1230 DTPSDEPTPSD--EPTPSDEPT--PSDEPTP--SETEEP--IPTDTPSD 1277
OY 707 ATKPEMTTAKKTTTERDLRTTPTTTAPKMTKETATTTTEKSTIATTTQVSTTTT 766
DB 1278 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSSETPPEEPIPTDPSDEPTPS 1337
OY 767 QDTPPKKITLKTTLAPKVTTPKKTITTELMNKKEETAKPKDRATNSKATTPKQK-XP 825
DB 1338 DEPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPT 1372
OY 826 TKAPKKTSTKPKKTPRVRKPKTTPPKKMTS-----TMBELNTPSRIA 870
DB 1373 SEPTPEPTPTTTP--TPTSTPTSGSGSGSGSGGGGGGGGTVPTSPPTPTPS-- 1424

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L21998; AAB95295.1; -  
DR EMBL: M74027; AAB59875.1; -  
DR EMBL: M94131; AAB59163.1; -  
DR EMBL: M94132; AAB59164.1; -  
DR MIM: 158370; -  
DR InterPro: IPR000359; Cys\_knot.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR002400; GF\_cys\_knot.  
DR InterPro: IPR001007; VMEC.  
DR InterPro: IPR001846; Vwd.  
DR Pfam: PF00007; Cys\_knot. 1.  
DR Pfam: PF00094; Vwd. 4.  
DR PRINTS: PR00438; GFCSKNOT.  
DR SMART: SM00214; VMC; 2.  
DR SMART: SM00011; VMC-def; 2.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; VMEC; 2.  
DR GlycoProtein: Repeat; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 5179  
FT DOMAIN 1401 1747  
FT REPEAT 1401 1416  
FT REPEAT 1417 1432  
FT REPEAT 1433 1448  
FT REPEAT 1449 1464  
FT REPEAT 1465 1471  
FT REPEAT 1472 1478  
FT REPEAT 1479 1494  
FT REPEAT 1495 1517  
FT REPEAT 1518 1533  
FT REPEAT 1534 1556  
FT REPEAT 1557 1572  
FT REPEAT 1573 1596  
FT REPEAT 1597 1612  
FT REPEAT 1613 1635  
FT REPEAT 1636 1651  
FT REPEAT 1652 1675  
FT REPEAT 1676 1693  
FT REPEAT 1694 1715  
FT REPEAT 1716 1731  
FT REPEAT 1732 1747  
FT DOMAIN 1748 1763  
FT DOMAIN 1764 1779  
FT DOMAIN 1780 1795  
FT DOMAIN 1796 1811  
FT DOMAIN 1812 1827  
FT DOMAIN 1828 1843  
FT DOMAIN 1844 1859  
FT DOMAIN 1860 1875  
FT DOMAIN 1876 1891  
FT DOMAIN 1892 1907  
FT DOMAIN 1908 1923  
FT DOMAIN 1924 1939  
FT DOMAIN 1940 1955  
FT DOMAIN 1956 1971  
FT DOMAIN 1972 1987  
FT DOMAIN 1988 2003  
FT DOMAIN 2004 2019  
FT DOMAIN 2020 2035  
FT DOMAIN 2036 2051  
FT DOMAIN 2052 2067  
FT DOMAIN 2068 2083  
FT DOMAIN 2084 2099  
FT DOMAIN 2100 2115  
FT DOMAIN 2116 2131  
FT DOMAIN 2132 2147  
FT DOMAIN 2148 2163  
FT DOMAIN 2164 2179  
FT DOMAIN 2180 2195  
FT DOMAIN 2196 2211  
FT DOMAIN 2212 2227  
FT DOMAIN 2228 2243  
FT DOMAIN 2244 2259  
FT DOMAIN 2260 2275  
FT DOMAIN 2276 2291  
FT DOMAIN 2292 2307  
FT DOMAIN 2308 2323  
FT DOMAIN 2324 2339  
FT DOMAIN 2340 2355  
FT DOMAIN 2356 2371  
FT DOMAIN 2372 2387  
FT DOMAIN 2388 2403  
FT DOMAIN 2404 2419  
FT DOMAIN 2420 2435  
FT DOMAIN 2436 2451  
FT DOMAIN 2452 2467  
FT DOMAIN 2468 2483  
FT DOMAIN 2484 2499  
FT DOMAIN 2500 2515  
FT DOMAIN 2516 2531  
FT DOMAIN 2532 2547  
FT DOMAIN 2548 2563  
FT DOMAIN 2564 2579  
FT DOMAIN 2580 2595  
FT DOMAIN 2596 2611  
FT DOMAIN 2612 2627  
FT DOMAIN 2628 2643  
FT DOMAIN 2644 2659  
FT DOMAIN 2660 2675  
FT DOMAIN 2676 2691  
FT DOMAIN 2692 2707  
FT DOMAIN 2708 2723  
FT DOMAIN 2724 2739  
FT DOMAIN 2740 2755  
FT DOMAIN 2756 2771  
FT DOMAIN 2772 2787  
FT DOMAIN 2788 2803  
FT DOMAIN 2804 2819  
FT DOMAIN 2820 2835  
FT DOMAIN 2836 2851  
FT DOMAIN 2852 2867  
FT DOMAIN 2868 2883  
FT DOMAIN 2884 2899  
FT DOMAIN 2900 2915  
FT DOMAIN 2916 2931  
FT DOMAIN 2932 2947  
FT DOMAIN 2948 2963  
FT DOMAIN 2964 2979  
FT DOMAIN 2980 2995  
FT DOMAIN 2996 3011  
FT DOMAIN 3012 3027  
FT DOMAIN 3028 3043  
FT DOMAIN 3044 3059  
FT DOMAIN 3060 3075  
FT DOMAIN 3076 3091  
FT DOMAIN 3092 3107  
FT DOMAIN 3108 3123  
FT DOMAIN 3124 3139  
FT DOMAIN 3140 3155  
FT DOMAIN 3156 3171  
FT DOMAIN 3172 3187  
FT DOMAIN 3188 3203  
FT DOMAIN 3204 3219  
FT DOMAIN 3220 3235  
FT DOMAIN 3236 3251  
FT DOMAIN 3252 3267  
FT DOMAIN 3268 3283  
FT DOMAIN 3284 3299  
FT DOMAIN 3300 3315  
FT DOMAIN 3316 3331  
FT DOMAIN 3332 3347  
FT DOMAIN 3348 3363  
FT DOMAIN 3364 3379  
FT DOMAIN 3380 3395  
FT DOMAIN 3396 3411  
FT DOMAIN 3412 3427  
FT DOMAIN 3428 3443  
FT DOMAIN 3444 3459  
FT DOMAIN 3460 3475  
FT DOMAIN 3476 3491  
FT DOMAIN 3492 3507  
FT DOMAIN 3508 3523  
FT DOMAIN 3524 3539  
FT DOMAIN 3540 3555  
FT DOMAIN 3556 3571  
FT DOMAIN 3572 3587  
FT DOMAIN 3588 3603  
FT DOMAIN 3604 3619  
FT DOMAIN 3620 3635  
FT DOMAIN 3636 3651  
FT DOMAIN 3652 3667  
FT DOMAIN 3668 3683  
FT DOMAIN 3684 3699  
FT DOMAIN 3700 3715  
FT DOMAIN 3716 3731  
FT DOMAIN 3732 3747  
FT DOMAIN 3748 3763  
FT DOMAIN 3764 3779  
FT DOMAIN 3780 3795  
FT DOMAIN 3796 3811  
FT DOMAIN 3812 3827  
FT DOMAIN 3828 3843  
FT DOMAIN 3844 3859  
FT DOMAIN 3860 3875  
FT DOMAIN 3876 3891  
FT DOMAIN 3892 3907  
FT DOMAIN 3908 3923  
FT DOMAIN 3924 3939  
FT DOMAIN 3940 3955  
FT DOMAIN 3956 3971  
FT DOMAIN 3972 3987  
FT DOMAIN 3988 4003  
FT DOMAIN 4004 4019  
FT DOMAIN 4020 4035  
FT DOMAIN 4036 4051  
FT DOMAIN 4052 4067  
FT DOMAIN 4068 4083  
FT DOMAIN 4084 4099  
FT DOMAIN 4100 4115  
FT DOMAIN 4116 4131  
FT DOMAIN 4132 4147  
FT DOMAIN 4148 4163  
FT DOMAIN 4164 4179  
FT DOMAIN 4180 4195  
FT DOMAIN 4196 4211  
FT DOMAIN 4212 4227  
FT DOMAIN 4228 4243  
FT DOMAIN 4244 4259  
FT DOMAIN 4260 4275  
FT DOMAIN 4276 4291  
FT DOMAIN 4292 4307  
FT DOMAIN 4308 4323  
FT DOMAIN 4324 4339  
FT DOMAIN 4340 4355  
FT DOMAIN 4356 4371  
FT DOMAIN 4372 4387  
FT DOMAIN 4388 4403  
FT DOMAIN 4404 4419  
FT DOMAIN 4420 4435  
FT DOMAIN 4436 4451  
FT DOMAIN 4452 4467  
FT DOMAIN 4468 4483  
FT DOMAIN 4484 4499  
FT DOMAIN 4500 4515  
FT DOMAIN 4516 4531  
FT DOMAIN 4532 4547  
FT DOMAIN 4548 4563  
FT DOMAIN 4564 4579  
FT DOMAIN 4580 4595  
FT DOMAIN 4596 4611  
FT DOMAIN 4612 4627  
FT DOMAIN 4628 4643  
FT DOMAIN 4644 4659  
FT DOMAIN 4660 4675  
FT DOMAIN 4676 4691  
FT DOMAIN 4692 4707  
FT DOMAIN 4708 4723  
FT DOMAIN 4724 4739  
FT DOMAIN 4740 4755  
FT DOMAIN 4756 4771  
FT DOMAIN 4772 4787  
FT DOMAIN 4788 4803  
FT DOMAIN 4804 4819  
FT DOMAIN 4820 4835  
FT DOMAIN 4836 4851  
FT DOMAIN 4852 4867  
FT DOMAIN 4868 4883  
FT DOMAIN 4884 4899  
FT DOMAIN 4900 4915  
FT DOMAIN 4916 4931  
FT DOMAIN 4932 4947  
FT DOMAIN 4948 4963  
FT DOMAIN 4964 4979  
FT DOMAIN 4980 4995  
FT DOMAIN 4996 5011  
FT DOMAIN 5012 5027  
FT DOMAIN 5028 5043  
FT DOMAIN 5044 5059  
FT DOMAIN 5060 5075  
FT DOMAIN 5076 5091  
FT DOMAIN 5092 5107  
FT DOMAIN 5108 5123  
FT DOMAIN 5124 5139  
FT DOMAIN 5140 5155  
FT DOMAIN 5156 5171  
FT DOMAIN 5172 5187  
FT DOMAIN 5188 5203  
FT DOMAIN 5204 5219  
FT DOMAIN 5220 5235  
FT DOMAIN 5236 5251  
FT DOMAIN 5252 5267  
FT DOMAIN 5268 5283  
FT DOMAIN 5284 5299  
FT DOMAIN 5300 5315  
FT DOMAIN 5316 5331  
FT DOMAIN 5332 5347  
FT DOMAIN 5348 5363  
FT DOMAIN 5364 5379  
FT DOMAIN 5380 5395  
FT DOMAIN 5396 5411  
FT DOMAIN 5412 5427  
FT DOMAIN 5428 5443  
FT DOMAIN 5444 5459  
FT DOMAIN 5460 5475  
FT DOMAIN 5476 5491  
FT DOMAIN 5492 5507  
FT DOMAIN 5508 5523  
FT DOMAIN 5524 5539  
FT DOMAIN 5540 5555  
FT DOMAIN 5556 5571  
FT DOMAIN 5572 5587  
FT DOMAIN 5588 5603  
FT DOMAIN 5604 5619  
FT DOMAIN 5620 5635  
FT DOMAIN 5636 5651  
FT DOMAIN 5652 5667  
FT DOMAIN 5668 5683  
FT DOMAIN 5684 5699  
FT DOMAIN 5700 5715  
FT DOMAIN 5716 5731  
FT DOMAIN 5732 5747  
FT DOMAIN 5748 5763  
FT DOMAIN 5764 5779  
FT DOMAIN 5780 5795  
FT DOMAIN 5796 5811  
FT DOMAIN 5812 5827  
FT DOMAIN 5828 5843  
FT DOMAIN 5844 5859  
FT DOMAIN 5860 5875  
FT DOMAIN 5876 5891  
FT DOMAIN 5892 5907  
FT DOMAIN 5908 5923  
FT DOMAIN 5924 5939  
FT DOMAIN 5940 5955  
FT DOMAIN 5956 5971  
FT DOMAIN 5972 5987  
FT DOMAIN 5988 6003  
FT DOMAIN 6004 6019  
FT DOMAIN 6020 6035  
FT DOMAIN 6036 6051  
FT DOMAIN 6052 6067  
FT DOMAIN 6068 6083  
FT DOMAIN 6084 6099  
FT DOMAIN 6100 6115  
FT DOMAIN 6116 6131  
FT DOMAIN 6132 6147  
FT DOMAIN 6148 6163  
FT DOMAIN 6164 6179  
FT DOMAIN 6180 6195  
FT DOMAIN 6196 6211  
FT DOMAIN 6212 6227  
FT DOMAIN 6228 6243  
FT DOMAIN 6244 6259  
FT DOMAIN 6260 6275  
FT DOMAIN 6276 6291  
FT DOMAIN 6292 6307  
FT DOMAIN 6308 6323  
FT DOMAIN 6324 6339  
FT DOMAIN 6340 6355  
FT DOMAIN 6356 6371  
FT DOMAIN 6372 6387  
FT DOMAIN 6388 6403  
FT DOMAIN 6404 6419  
FT DOMAIN 6420 6435  
FT DOMAIN 6436 6451  
FT DOMAIN 6452 6467  
FT DOMAIN 6468 6483  
FT DOMAIN 6484 6499  
FT DOMAIN 6500 6515  
FT DOMAIN 6516 6531  
FT DOMAIN 6532 6547  
FT DOMAIN 6548 6563  
FT DOMAIN 6564 6579  
FT DOMAIN 6580 6595  
FT DOMAIN 6596 6611  
FT DOMAIN 6612 6627  
FT DOMAIN 6628 6643  
FT DOMAIN 6644 6659  
FT DOMAIN 6660 6675  
FT DOMAIN 6676 6691  
FT DOMAIN 6692 6707  
FT DOMAIN 6708 6723  
FT DOMAIN 6724 6739  
FT DOMAIN 6740 6755  
FT DOMAIN 6756 6771  
FT DOMAIN 6772 6787  
FT DOMAIN 6788 6803  
FT DOMAIN 6804 6819  
FT DOMAIN 6820 6835  
FT DOMAIN 6836 6851  
FT DOMAIN 6852 6867  
FT DOMAIN 6868 6883  
FT DOMAIN 6884 6899  
FT DOMAIN 6900 6915  
FT DOMAIN 6916 6931  
FT DOMAIN 6932 6947  
FT DOMAIN 6948 6963  
FT DOMAIN 6964 6979  
FT DOMAIN 6980 6995  
FT DOMAIN 6996 7011  
FT DOMAIN 7012 7027  
FT DOMAIN 7028 7043  
FT DOMAIN 7044 7059  
FT DOMAIN 7060 7075  
FT DOMAIN 7076 7091  
FT DOMAIN 7092 7107  
FT DOMAIN 7108 7123  
FT DOMAIN 7124 7139  
FT DOMAIN 7140 7155  
FT DOMAIN 7156 7171  
FT DOMAIN 7172 7187  
FT DOMAIN 7188 7203  
FT DOMAIN 7204 7219  
FT DOMAIN 7220 7235  
FT DOMAIN 7236 7251  
FT DOMAIN 7252 7267  
FT DOMAIN 7268 7283  
FT DOMAIN 7284 7299  
FT DOMAIN 7300 7315  
FT DOMAIN 7316 7331  
FT DOMAIN 7332 7347  
FT DOMAIN 7348 7363  
FT DOMAIN 7364 7379  
FT DOMAIN 7380 7395  
FT DOMAIN 7396 7411  
FT DOMAIN 7412 7427  
FT DOMAIN 7428 7443  
FT DOMAIN 7444 7459  
FT DOMAIN 7460 7475  
FT DOMAIN 7476 7491  
FT DOMAIN 7492 7507  
FT DOMAIN 7508 7523  
FT DOMAIN 7524 7539  
FT DOMAIN 7540 7555  
FT DOMAIN 7556 7571  
FT DOMAIN 7572 7587  
FT DOMAIN 7588 7603  
FT DOMAIN 7604 7619  
FT DOMAIN 7620 7635  
FT DOMAIN 7636 7651  
FT DOMAIN 7652 7667  
FT DOMAIN 7668 7683  
FT DOMAIN 7684 7699  
FT DOMAIN 7700 7715  
FT DOMAIN 7716 7731  
FT DOMAIN 7732 7747  
FT DOMAIN 7748 7763  
FT DOMAIN 7764 7779  
FT DOMAIN 7780 7795  
FT DOMAIN 7796 7811  
FT DOMAIN 7812 7827  
FT DOMAIN 7828 7843  
FT DOMAIN 7844 7859  
FT DOMAIN 7860 7875  
FT DOMAIN 7876 7891  
FT DOMAIN 7892 7907  
FT DOMAIN 7908 7923  
FT DOMAIN 7924 7939  
FT DOMAIN 7940 7955  
FT DOMAIN 7956 7971  
FT DOMAIN 7972 7987  
FT DOMAIN 7988 8003  
FT DOMAIN 8004 8019  
FT DOMAIN 8020 8035  
FT DOMAIN 8036 8051  
FT DOMAIN 8052 8067  
FT DOMAIN 8068 8083  
FT DOMAIN 8084 8099  
FT DOMAIN 8100 8115  
FT DOMAIN 8116 8131  
FT DOMAIN 8132 8147  
FT DOMAIN 8148 8163  
FT DOMAIN 8164 8179  
FT DOMAIN 8180 8195  
FT DOMAIN 8196 8211  
FT DOMAIN 8212 8227  
FT DOMAIN 8228 8243  
FT DOMAIN 8244 8259  
FT DOMAIN 8260 8275  
FT DOMAIN 8276 8291  
FT DOMAIN 8292 8307  
FT DOMAIN 8308 8323  
FT DOMAIN 8324 8339  
FT DOMAIN 8340 8355  
FT DOMAIN 8356 8371  
FT DOMAIN 8372 8387  
FT DOMAIN 8388 8403  
FT DOMAIN 8404 8419  
FT DOMAIN 8420 8435  
FT DOMAIN 8436 8451  
FT DOMAIN 8452 8467  
FT DOMAIN 8468 8483  
FT DOMAIN 8484 8499  
FT DOMAIN 8500 8515  
FT DOMAIN 8516 8531  
FT DOMAIN 8532 8547  
FT DOMAIN 8548 8563  
FT DOMAIN 8564 8579  
FT DOMAIN 8580 8595  
FT DOMAIN 8596 8611  
FT DOMAIN 8612 8627  
FT DOMAIN 8628 8643  
FT DOMAIN 8644 8659  
FT DOMAIN 8660 8675  
FT DOMAIN 8676 8691  
FT DOMAIN 8692 8707  
FT DOMAIN 8708 8723  
FT DOMAIN 8724 8739  
FT DOMAIN 8740 8755  
FT DOMAIN 8756 8771  
FT DOMAIN 8772 8787  
FT DOMAIN 8788 8803  
FT DOMAIN 8804 8819  
FT DOMAIN 8820 8835  
FT DOMAIN 8836 8851  
FT DOMAIN 8852 8867  
FT DOMAIN 8868 8883  
FT DOMAIN 8884 8899  
FT DOMAIN 8900 8915  
FT DOMAIN 8916 8931  
FT DOMAIN 8932 8947  
FT DOMAIN 8948 8963  
FT DOMAIN 8964 8979  
FT DOMAIN 8980 8995  
FT DOMAIN 8996 9011  
FT DOMAIN 9012 9027  
FT DOMAIN 9028 9043  
FT DOMAIN 9044 9059  
FT DOMAIN 9060 9075  
FT DOMAIN 9076 9091  
FT DOMAIN 9092 9107  
FT DOMAIN 9108 9123  
FT DOMAIN 9124 9139  
FT DOMAIN 9140 9155  
FT DOMAIN 9156 9171  
FT DOMAIN 9172 9187  
FT DOMAIN 9188 9203  
FT DOMAIN 9204 9219  
FT DOMAIN 9220 9235  
FT DOMAIN 9236 9251  
FT DOMAIN 9252 9267  
FT DOMAIN 9268 9283  
FT DOMAIN 9284 9299  
FT DOMAIN 9300 9315  
FT DOMAIN 9316 9331  
FT DOMAIN 9332 9347  
FT DOMAIN 9348 9363  
FT DOMAIN 9364 9379  
FT DOMAIN 9380 9395  
FT DOMAIN 9396 9411  
FT DOMAIN 9412 9427  
FT DOMAIN 9428 9443  
FT DOMAIN 9444 9459  
FT DOMAIN 9460 9475  
FT DOMAIN 9476 9491  
FT DOMAIN 9492 9507  
FT DOMAIN 9508 9523  
FT DOMAIN 9524 9539  
FT DOMAIN 9540 9555  
FT DOMAIN 9556 9571  
FT DOMAIN 9572 9587  
FT DOMAIN 9588 9603  
FT DOMAIN 9604 9619  
FT DOMAIN 9620 9635  
FT DOMAIN 9636 9651  
FT DOMAIN 9652 9667  
FT DOMAIN 9668 9683  
FT DOMAIN 9684 9699  
FT DOMAIN 9700 9715  
FT DOMAIN 9716 9731  
FT DOMAIN 9732 9747  
FT DOMAIN 9748 9763  
FT DOMAIN 9764 9779  
FT DOMAIN 9780 9795  
FT DOMAIN 9796 9811  
FT DOMAIN 9812 9827  
FT DOMAIN 9828 9843  
FT DOMAIN 9844 9859  
FT DOMAIN 9860 9875  
FT DOMAIN 9876 9891  
FT DOMAIN 9892 9907  
FT DOMAIN 9908 9923  
FT DOMAIN 9924 9939  
FT DOMAIN 9940 9955  
FT DOMAIN 9956 9971  
FT DOMAIN 9972 9987  
FT DOMAIN 9988 10003  
FT DOMAIN 10004 10019  
FT DOMAIN 10020 10035  
FT DOMAIN 10036 10051  
FT DOMAIN 10052 10067  
FT DOMAIN 10068 10083  
FT DOMAIN 10084 10099  
FT DOMAIN 10100 10115  
FT DOMAIN 10116 10131  
FT DOMAIN 10132 10147  
FT DOMAIN 10148 10163  
FT DOMAIN 10164 10179  
FT DOMAIN 10180 10195  
FT DOMAIN 10196 10211  
FT DOMAIN 10212 10227  
FT DOMAIN 10228 10243  
FT DOMAIN 10244 10259  
FT DOMAIN 10260 10275  
FT DOMAIN 10276 10291  
FT DOMAIN 10292 10307  
FT DOMAIN 10308 10323  
FT DOMAIN 10324 10339  
FT DOMAIN 10340 10355  
FT DOMAIN 10356 10371  
FT DOMAIN 10372 10387  
FT DOMAIN 10388 10403  
FT DOMAIN 10404 10419  
FT DOMAIN 10420 10435  
FT DOMAIN 10436 10451  
FT DOMAIN 10452 10467  
FT DOMAIN 10468 10483  
FT DOMAIN 10484 10499  
FT DOMAIN 10500 10515  
FT DOMAIN 10516 10531  
FT DOMAIN 10532 10547  
FT DOMAIN 10548 10563  
FT DOMAIN 10564 10579  
FT DOMAIN 10580 10595  
FT DOMAIN 10596 10611  
FT DOMAIN 10612 10627  
FT DOMAIN 10628 10643  
FT DOMAIN 10644 10659  
FT DOMAIN 10660 10675  
FT DOMAIN 10676 10691  
FT DOMAIN 10692 10707  
FT DOMAIN 10708 10723  
FT DOMAIN 10724 10739  
FT DOMAIN 10740 10755  
FT DOMAIN 10756 10771  
FT DOMAIN 10772 10787  
FT DOMAIN 10788 10803  
FT DOMAIN 10804 10819  
FT DOMAIN 10820 10835  
FT DOMAIN 10836 10851  
FT DOMAIN 10852 10867  
FT DOMAIN 10868 10883  
FT DOMAIN 10884 10899  
FT DOMAIN 10900 10915  
FT DOMAIN 10916 10931  
FT DOMAIN 10932 10947  
FT DOMAIN 10948 10963  
FT DOMAIN 10964 10979  
FT DOMAIN 10980 10995  
FT DOMAIN 10996 11011  
FT DOMAIN 11012 11027  
FT DOMAIN 11028 11043  
FT DOMAIN 11044 11059  
FT DOMAIN 11060 11075  
FT DOMAIN 11076 11091  
FT DOMAIN 11092 11107  
FT DOMAIN 11108 11123  
FT DOMAIN 11124 11139  
FT DOMAIN 11140 11155  
FT DOMAIN 11156 11171  
FT DOMAIN 11172 11187  
FT DOMAIN 11188 11203  
FT DOMAIN 11204 11219  
FT DOMAIN 11220 11235  
FT DOMAIN 11236 11251  
FT DOMAIN 11252 11267  
FT DOMAIN 11268 11283  
FT DOMAIN 11284 11299  
FT DOMAIN 11300 11315  
FT DOMAIN 11316 11331  
FT DOMAIN 11332 11347  
FT DOMAIN 11348 11363  
FT DOMAIN 11364 11379  
FT DOMAIN 11380 11395  
FT DOMAIN 11396 11411  
FT DOMAIN 11412 11427  
FT DOMAIN 11428 11443  
FT DOMAIN 11444 11459  
FT DOMAIN 11460 11475  
FT DOMAIN 11476 11491  
FT DOMAIN 11492 11507  
FT DOMAIN 11508 11523  
FT DOMAIN 11524 11539  
FT DOMAIN 11540 11555  
FT DOMAIN 11556 11571  
FT DOMAIN 11572 11587  
FT DOMAIN 11588 11603  
FT DOMAIN 11604 11619  
FT DOMAIN 11620 11635  
FT DOMAIN 11636 11651  
FT DOMAIN 11652 11667  
FT DOMAIN 11668 11683  
FT DOMAIN 11684 11699  
FT DOMAIN 11700 11715  
FT DOMAIN 11716 11731  
FT DOMAIN 11732 11747  
FT DOMAIN 11748 11763  
FT DOMAIN 11764 11779  
FT DOMAIN 11780 11795  
FT DOMAIN 11796 11811  
FT DOMAIN 11812 11827  
FT DOMAIN 11828 11843  
FT DOMAIN 11844 11859  
FT DOMAIN 11860 11875  
FT DOMAIN 11876 11891  
FT DOMAIN 11892 11907  
FT DOMAIN 11908 11923  
FT DOMAIN 11924 11939  
FT DOMAIN 11940 11955  
FT DOMAIN 11956 11971  
FT DOMAIN 11972 11987  
FT DOMAIN 11988 12003  
FT DOMAIN 12004 12019  
FT DOMAIN 12020 12035  
FT DOMAIN 12036 12051  
FT DOMAIN 12052 12067  
FT DOMAIN 12068 12083  
FT DOMAIN 12084 12099  
FT DOMAIN 12100 12115  
FT DOMAIN 12116 12131  
FT DOMAIN 12132 12147  
FT DOMAIN 12148 12163  
FT DOMAIN 12164 12179  
FT DOMAIN 12180 12195  
FT DOMAIN 12196 12211  
FT DOMAIN 12212 12227  
FT DOMAIN 12228 12243  
FT DOMAIN 12244 12259  
FT DOMAIN 12260 12275  
FT DOMAIN 12276 12291  
FT DOMAIN 12292 12307  
FT DOMAIN 12308 12323  
FT DOMAIN 12324 12339  
FT DOMAIN 12340 12355  
FT DOMAIN 12356 12371  
FT DOMAIN 12372 12387  
FT DOMAIN 12388 12403  
FT DOMAIN 12404 12419  
FT DOMAIN 12420 12435  
FT DOMAIN 12436 12451  
FT DOMAIN 12452 12467  
FT DOMAIN 12468 12483  
FT DOMAIN 12484 12499  
FT DOMAIN 12500 12515  
FT DOMAIN 12516 12531  
FT DOMAIN 12532 12547  
FT DOMAIN 12548 12563  
FT DOMAIN 12564 12579  
FT DOMAIN 12580 12595  
FT DOMAIN 12596 12611  
FT DOMAIN 12612 12627  
FT DOMAIN 12628 12643  
FT DOMAIN 12644 12659  
FT DOMAIN 12660 12675  
FT DOMAIN 12676 12691  
FT DOMAIN 12692 12707  
FT DOMAIN 12708 12723  
FT DOMAIN 12724 12739  
FT DOMAIN 12740 12755  
FT DOMAIN 12756 12771  
FT DOMAIN 12772 12787  
FT DOMAIN 12788 12803  
FT DOMAIN 12804 12819  
FT DOMAIN 12820 12835  
FT DOMAIN 12836 12851  
FT DOMAIN 12852 12867  
FT DOMAIN 12868 12883  
FT DOMAIN 12884 12899  
FT DOMAIN 12900 12915  
FT DOMAIN 12916 12931  
FT DOMAIN 12932 12947  
FT DOMAIN 12948 12963  
FT DOMAIN 12964 12979  
FT DOMAIN 12980 12995  
FT DOMAIN 12996 13011  
FT DOMAIN 13012 13027  
FT DOMAIN 13028 13043  
FT DOMAIN 13044 13059  
FT DOMAIN 13060 13075  
FT DOMAIN 13076 13091  
FT DOMAIN 13092 13107  
FT DOMAIN 13108 13123  
FT DOMAIN 13124 13139  
FT DOMAIN 13140 13155  
FT DOMAIN 13156 13171  
FT DOMAIN 13172 13187  
FT DOMAIN 13188 13203  
FT DOMAIN 13204 13219  
FT DOMAIN 13220 13235  
FT DOMAIN 13236 13251  
FT DOMAIN 13252 13267  
FT DOMAIN 13268 13283  
FT DOMAIN 13284 13299  
FT DOMAIN 13300 13315  
FT DOMAIN 13316 13331  
FT DOMAIN 13332 13347  
FT DOMAIN 13348 13363  
FT DOMAIN 13364 13379  
FT DOMAIN 13380 13395  
FT DOMAIN 13396 13411  
FT DOMAIN 13412 13427  
FT DOMAIN 13428 13443  
FT DOMAIN 13444 13459  
FT DOMAIN 13460 13475  
FT DOMAIN 13476 13491  
FT DOMAIN 13492 13507  
FT DOMAIN 13508 13523  
FT DOMAIN 13524 13539  
FT DOMAIN 13540 13555  
FT DOMAIN 13556 13571  
FT DOMAIN 13572 13587  
FT DOMAIN 13588 13603  
FT DOMAIN 13604 13619  
FT DOMAIN 13620 13635  
FT DOMAIN 13636 13651  
FT DOMAIN 13652 13667  
FT DOMAIN 13668 13683  
FT DOMAIN 13684 13699  
FT DOMAIN 13700 13715  
FT DOMAIN 13716 13731  
FT DOMAIN 13732 13747  
FT DOMAIN 13748 13763  
FT DOMAIN 13764 13779  
FT DOMAIN 13780 13795  
FT DOMAIN 13796 13811  
FT DOMAIN 13812 13827  
FT DOMAIN 13828 13843  
FT DOMAIN 13844 13859  
FT DOMAIN 13860 13875  
FT DOMAIN 13876 13891  
FT DOMAIN 13892 13907  
FT DOMAIN 13908 13923  
FT DOMAIN 13924 13939  
FT DOMAIN 13940 13955  
FT DOMAIN 13956 13971  
FT DOMAIN 13972 13987  
FT DOMAIN 13988 14003  
FT DOMAIN 14004 14019  
FT DOMAIN 14020 14035  
FT DOMAIN 14036 14051  
FT DOMAIN 14052 14067  
FT DOMAIN 14068 14083  
FT

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:35 ; Search time 47.4 Seconds  
(without alignments)  
727.882 Million cell updates/sec

Title: US-09-556-246-1-COPY\_200\_1140  
5011  
Perfect score: 1 VKOKKKRRTKKKPPKPPV.....DMDYLPVPNGIINPMLS 941  
Sequence:

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_39.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1168	23.3	5179	1 MUC2_HUMAN	002817 homo sapien
2	942	18.8	1664	1 SLPL_CLOTT	006852 clostridium
3	762.3	15.2	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	13.0	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	11.1	875	1 FPL_MYTD	Q25460 mytilus edu
6	555.5	11.1	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	551	11.0	620	1 EXTN_TOBAC	P13963 nicotiana t
8	533	10.6	1087	1 NEH_MOUSE	P19246 mus musculu
9	530.5	10.6	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	10.6	865	1 CPN_DROME	Q02910 drosophila
11	518.5	10.3	872	1 FPL_MYTCO	Q25434 mytilus cor
12	509.5	10.2	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	10.0	1970	1 RPB1_HUMAN	P24928 homo sapien
14	498.5	9.9	1970	1 RPB1_MOUSE	P08775 mus musculu
15	497	9.9	831	1 NEH_RAT	P16884 rattus norv
16	493.5	9.8	467	1 RPB1_CRIGR	P11443 placetulus
17	488.5	9.7	826	1 SSP2_PLAYO	Q01443 plasmodium
18	488	9.7	1020	1 NEH_HUMAN	P12036 homo sapien
19	475.5	9.5	267	1 EXTN_MAIZE	P14918 zea mays (m
20	468.5	9.3	634	1 HNP1_CANAL	P46593 candida alb
21	467.5	9.3	5376	1 ZAN_MOUSE	Q08799 mus musculu
22	467	9.3	817	1 VRP1_YEAST	P37370 saccharomyc
23	454	9.1	1161	1 Y9P_YEAST	P47179 saccharomyc
24	452.5	9.0	2142	1 BAT2_HUMAN	P48634 homo sapien
25	448.5	9.0	670	1 VG50_HSV1	Q00130 ictalrid h
26	448	8.9	797	1 VGIX_HSVB	P28968 equine hept
27	442.5	8.8	751	1 FPL_MYTGA	Q27409 mytilus gal
28	439.5	8.8	1083	1 T2D3_HUMAN	Q00268 homo sapien
29	432.5	8.6	439	1 XP2_XENLA	P17437 xenopus lae
30	432	8.5	3164	1 TEGU_HSV1	P10220 herpes simp
31	426	8.5	2715	1 TRX2_HUMAN	Q9um66 homo sapien
32	424.5	8.5	1125	1 MAP4_MOUSE	P27546 mus musculu
33	424	8.5	307	1 SG53_DROME	P02840 drosophila

34	424	8.5	1251	1 Y0U3_CABEL	Q09550 caenorhabdi
35	422.5	8.4	1794	1 YAVI_SCHPO	Q10172 schizosacch
36	422.5	8.4	3421	1 TEGU_HSVB	P28955 equine hept
37	419.5	8.4	2476	1 ZAN_PTG	Q28983 sus scrofa
38	419	8.4	1185	1 DRPL_HUMAN	P54259 homo sapien
39	415	8.3	1229	1 N121_HUMAN	Q9Y293 homo sapien
40	411.5	8.2	1183	1 VGP3_EBV	P03200 Epstein-bar
41	410	8.2	1183	1 DRPL_RAT	P54258 rattus norv
42	406	8.1	2774	1 MAPA_RAT	P34926 rattus norv
43	406	8.1	3256	1 K167_HUMAN	P46013 homo sapien
44	405	8.1	1411	1 TCOE_HUMAN	Q13428 homo sapien
45	400.5	8.0	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r

## ALIGNMENTS

RESULT	1	STANDARD:	PRT: 5179 AA.
MUC2_HUMAN			
ID	MUC2_HUMAN	014878;	
AC	002817; 014878;		
DT	01-JUN-1994 (Rel. 29, Created)		
DR	01-NOV-1997 (Rel. 35, Last sequence update)		
DE	20-AUG-2001 (Rel. 40, Last annotation update)		
DI	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).		
GN	MUC2 OR SMC.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxId=9606;		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Intestine;		
RX	MEDLINE=94132002; PubMed=8300571;		
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;		
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.		
RT	Identification of the amino terminus and overall sequence similarity		
RT	to prepro-von Willebrand factor.";		
RL	J. Biol. Chem. 269:2440-2446(1994).		
RN	[2].		
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.		
RC	TISSUE=Colon;		
RX	MEDLINE=93016075; PubMed=1400449;		
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-W., Lagace R.E.,		
RA	Kim Y.S.;		
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located		
RT	both upstream and downstream of its central repetitive region.";		
RT	J. Biol. Chem. 267:21375-21383(1992).		
RN	[3].		
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.		
RX	MEDLINE=91358717; PubMed=1865763;		
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,		
RA	Petersen G.M., Kim Y.S.;		
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays		
RT	and polymorphism.";		
RL	J. Clin. Invest. 88:1005-1013(1991).		
CC	-1- FUNCTION: CONTS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND		
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A		
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTION		
CC	AGENTS AT MUCOSAL SURFACES.		
CC	-1- SUBUNIT: MULTIMERIC.		
CC	-1- SUBCELLULAR LOCATION: SECRETED.		
CC	-1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,		
CC	BRONCHUS, CERVIX AND GALL BLADDER.		
CC	-1- PPM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR		
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).		
CC	-1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND		
CC	VARIES AMONG DIFFERENT ALLELES.		
CC	-1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT		
CC	OF SIKKORUM HEMOCYTIN.		
CC	-1- SIMILARITY: CONTAINS 2 WPC DOMAINS.		
CC	-1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).		

```
QY 117 TSAKDLAPTSVLAKEPTPKAETTTGPALETTKEPTPTTKE-----PASTTKEPTPTTI 172
Db 171 EASKKKQVIVK----PKQAVVKAAGLASNMNGKTADSSSESDSPAKKTAATKTPTPT- 225
QY 173 KSAPTTKEPAPTTTSAPTTPKEPAPTTKEPAPTTKEPAPTT--TKEPAPTT-----T 226
Db 226 --KPAATAKPOAKKTAGKSSREDSDEQKTAKSCKPDVYSAVPPPTSVSKKT 283
QY 227 KSAPTTKEPAPT-----TPKKPAPTTKEPAPTT---KEPTPTTKEPAPTT 272
Db 284 LSQPGTKAKPSSSDSSDSEBOPAKKATVPAAKAAASAPKAPLAKKATSTSESDSS 343
QY 273 KEPAPTTKEPAPTAPKKAPATTPKEPAPTT---KEPA-----TT 311
Db 344 EDEKSSVKGIVKAPKK-APAAP--DAKSTPVAAKKSAKAPAKKSSSDSDSSNEETT 400
QY 312 TKEPSPTTKEPAPTTTSAPTTPKEPAPTTKEPAPTTKEPSPTTKEPAPTTKEPA 370
Db 401 TKPAKKTTPAKSAATPTSKTPTNGK--ATPTSKTPAKGTPKTSTAKKSDSSSDSS 457
QY 371 --PTTPKKPAPTTKEPAPT--TPKEPAPTTTPKKPAAPAPKEPA--PTTPKETA----- 418
Db 458 SDEETTTTPAKATTPAKSAATPTSKTP---TNSKATPTSKTTPAKGTPKTSAKKDSS 514
QY 419 -----PTTPKKLTPTTPKLAPTTPEKPAPTTPEELAPTEPEPTPTTPEEPAPTTP--KA 472
Db 515 SDSSSDSSDEKKTTPA--KRAKTTTPAKPA-----AKTTTPAKP-----AKTTTPAKP 558
QY 473 AAPNTPKPAPT---TPKEPAPTTTPKEPAPTTKEPAPTTTP--KGTAPTTTPKEPAPTTPK 527
Db 559 AKSTTPGKQVPTKSSSDSSDSSDEKKSAPKPAVKTPPGKATS-----KPVVAS 611
QY 528 KPAKELAPTTTKEPTSTTSKPAPTTPKGAPTTPKEPAPTT--PKPAPTTPKGAPT 585
Db 612 KPVPK-----KASSSDSSSEETTKTKPLTKLSPAVKTLPPKKAESSSDSSSDSD 665
QY 586 TLKEPAPTTPKKAPKELAPTTTGP-----TSTTSKPAPTTPKETAPT--TPKEP 635
Db 666 SEK-----TKPAKPPAKSATPVNTKAPAKONKASKASCSDSSSEEGSKQPTGKSPA 722
QY 636 APTPPKK-PAPTTPETPTPTTSEVSTPTTKE-PTTIHKSPESTPELSAEPPTKALENS 693
Db 723 ATAPPKNPNVAVNKDSSSSSDSGDDEKQKQAAAKADVQGAKAAPTPPKKAASS 782
QY 694 KPEGVPPTTKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETATTEKTESK 753
Db 783 SSE---DSSDEDEVSKAKKTWTAVSKSPV---TTPKAVPAKAKSSSESSESDSEKQGG 835
QY 754 ITATTOVSTTODTTPFKTTTLKTTTLAKVTTTK--KTITTEIMNKPEETA-KPKD 810
Db 836 KNTSTTKIANST-----PKAAAECSSESSSEDEGKANGTSGKRR 877
QY 811 RATNS---KATTPPKQPTKAPKKPTSTKPKPTPRVRKPKPTTPTP--RKMTSTWPELNPT 866
Db 878 ESTGNACEAVTPE-----NKKLAKAKSPNTPKYNKKEKLNTPFRRYVEEDIEINP- 928
QY 867 SRIAE 871
Db 929 -RMAD 932
```

Search completed: April 26, 2002, 16:18:10  
Job time: 280 sec









Query Match	13.3%	Score 666.5;	DB 1;	Length 1344;
Best Local Similarity	28.48%;	Pred. No. 1.9e 21;		
Matches 302;	Conservative 70;	Mismatches 432;	Indels 261;	Gaps 57;

Db	723	PAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGS	---T	780
Qy	651	PAPTVSEVSTPTPTTKPTT	-----IHKSPDETPELSAEPKALEMSFKRECVPTTK	703
Db	781	APPAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGST	840	
Qy	704	TPAA-----TKREMTTAKOK	---TTERDLRTPTT-----TAAPKITKETAT	745
Db	841	APPAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAAP	---DTRPA	896
Qy	746	TEKTESKITATTOOVSTTTODTTP	---FKITTLKTTTLPAKVVTKTKKIIITTT---EIMNK	801
Db	897	PGSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPA	956	
Qy	802	PESTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPA	956	
Db	957	PGSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPA	956	
Qy	862	ELNPT-SKILAEMLQTTTRPQTP	---NSKLEVEYNPKSEDAGAEG	903
Db	1012	DTTPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPAAGVSTAPRAHVGTSAADTRPA	1056	

```

RESULT      10
t25697
Hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: t25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: t25697
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FOLU>
A:Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSEDB:GN00028; CESP:F16F9.2
C:Experimental source: Strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match      13.2%; Score 659.5; DB 2; Length 1229;
Best Local Similarity 29.0%; Pred. No. 3.5e-21;
Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52

QY      33 KVTT-PDSTNQ---HNKVSIPKLTFA-KPINPRSLPNSDQ-----SKESLTVN 80
Db      119 KVTSTDASTNAPITGKDSITPEITIGIVINSKSESVTDMSTREFSTLSPTTELLTS 178

QY      81 KETVETKEITTNKQSTODGKEKTTSAKEIOSIEKTSAKDIAPPSKVLAKEPRAKETT 140
Db      179 PETLVSDDSSNST-EQSPDNTEELASPEMNTTTEATTTSEVPSVSTLAS--EDETTT 234

QY      141 KGPALTTPKEEPTPTTPKBPASTTPKEEPTPTTTSKASPTTPKBPAPTTSAPTTPKEBAPT 200
Db      235 TAIASEST-----TTVIAEVSSTTEEPPTT-----AESTTKKSTT-----KAPA 272

QY      201 TTKBPAPTTPKBPAPTTTTPKBPAPTTTTSKAPPTTPKBPAPTPP---KKBPAPTTPKBPAPT 256
Db      273 TTEETPTPTTTE--VTTTEALSTTTSSETSTSEK---PTTPIDINKINGPATGK---DET 324

QY      257 PKBP-TPTTPKEPAPTTPKEPAPTTPKBPAPPAKKBPAPPTTPKEPAPTTPKBPAPTTTKE- 314
Db      325 THFPVGTGTPR--EDTATETPFAKSEDKMLSKAALETIOQTTEVNT-DGEKETTKNV 381

QY      315 -----PSPTTPKEPAPTTTTSAP-----TTTKEBPAP---TTSKAPPTTPKEPSPTTTPK 359
Db      382 SIETPTTVPLVETTSSTTSKSGDGHFTLLKLVTTADSSTESATTVKPFNEETTPK 441

```

```

Db      338 KRAPPPEQQLKRAAAAAAPITGELKPAATPHGSPRANSHTVTVTPVNPAAATVP- 396
      :||| : : : :||| : : : : : : : : : : : : : : : : : : : : : :
Qy      251 EAPPTTPEPTT-----PKEPAPTTKEPAPTKEPAPTAKKPAPTPKBPAPTPE 306
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      397 -TAGAVPAKSTGTTAAAOQDPV---KAAPVTPPSPQOAVPRAATAA-----APVTPOQ 448
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      307 P---APTTKEPSPPTTKEPAPTTPKSAPTTKEPAP---TTTKSAPTTKEPSPPTTKE 360
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      449 PTKATATTNNATPPQPIPKAATTTTATPVTPQPIPKAGTDAAPPAPKAPSDGRAT 508
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      361 P---APTTKEPAPTTPKAPAPTTPKEAP-----TTTKEAPAP---TTTKEAPAP 406
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      509 PCVPNAADPOKPPPTPOSVPASVTEPKDPQRAAPPSPNEAPVPVSPNLSPLPTIP 568
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      407 K-----EPAPTTPK---ETAPTTPKKLPTTPEKLAFTPEKAPPTPEELA----- 450
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      569 KVPPLMALTPQVTAOMTQOLAATKPSIVPKASPK-ALMTTPPPPPGLPRAALAAKLLG 627
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      451 -PTTP-----EPPPTTP---EPPAPTTP-----KAAPNTPE----- 480
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      628 LSSPVASAMHAKVTPRPAPSPVMAASPAISGPDAAVALATNAASGAKPEAGNG 687
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      481 ----PAPTTKEPAPTTPKEPAPTTPKEAPT-----TPKGAAPT----- 516
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      688 TLMAPMGAANTOMAPAGAAQATPMGAATHVSPMGAGATQMSPTGANTHMSFICA 747
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      517 -----TLKEPAPTTPKKPAKELAPTTTKEP-----TSTSDKPAPTTP 555
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      748 GGAATSPMGANTOMSPMGATTTQSPMGAAATQPSMGAAATQVATASAGTMOVS 807
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      556 KG--TAPTTKEPAPTTPKEPAPTTPKGAAPTTLKEP--APTTPKPAKELAPTTTKEP 611
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      808 MGAATPQATPQVGAATTP--QPSPM---GAATLMSPMGAATTPQ---PSPMGAVTTQPP 859
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      612 ----TSTSDKP--APTTPKEP---APTTPKEP---APTTPKPK--APTTPPTTP----- 652
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      860 PMAATNTQPPMAASTPOSTPMGATTTQSPMGATTTQSPMGASTPPAPPTVAGSPT 919
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      653 PTTSEVSTPTTKEPTTHKSPDESTEP--SAEPTPKALENSPKPEGVPTTTPAATPK 710
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      920 PPPPIPPSPQATSPQMSKSPPPDPKAPSAQAQTSFPAHVANASPGV--TAVSPA---P 975
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      711 EMTTTAKDKTERDILRTPEPTTAAPKMT--KEATTEKTESKINTVTTQVSTTQDT 769
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      976 IGVTEASPSADGARLSGPPAATDGPASPAATADVTEAATD--VTAATAVPA-----EA 1029
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      770 TPKKITLKTTLAPKVTYTKTITTTTETINMKPEETAKPKDRATNSKATTPKPKPTKAP 829
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      1030 AP-----TKAKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSNPASPA--- 1077
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      830 KKPTSTKKPTKPTKPTKPTTPPKKKTSTPELNPSTRIAEMLQTTTPNQTTPNKL 889
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      1078 -----PAVGDGQOQMTGPAQASVPP-----VTEAAVQ----- 1104
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Qy      890 EVNPKSEDAGAGE 904
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :
Db      1105 EAAAAAAGAAERE 1119
      :||| : : : :||| : : : :||| : : : :||| : : : :||| : : : :||| :

```

## RESULT 9

A35175  
 mucin 1 precursor, repetitive splice form A [validated] - human  
 N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; epistaltin  
 nceatic mucin; polymorphic epithelial mucin (PEM)  
 N:Contents: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor,  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Apr-2000 #sequence, revision 20-Apr-2000 #text, change 02-Jun-2000  
 C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218; S51  
 R;Ligtenberg, M.J.L.; Vos, H.L.; Gemtssen, A.M.C.; Hilkens, J.  
 J. Biol. Chem. 265, 5573-5578, 1990  
 A:Title: Epistaltin, a carcinoma-associated mucin, is generated by a polymorphic gene end  
 A:Reference number: A35175; MUID:90202794

A:Accession: A35175  
 A:Molecule type: mRNA  
 A:Residues: 1-952,1033-1344 <LIG1>  
 A:Cross-references: GB:M32739; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124;  
 A:Experimental source: splice form A  
 A:Note: Genbank entries HUMEPIS1 and HUMEPIS1A present only the amino-and carboxyl  
 A:Accession: B35175  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-952,1033-1344 <LIG2>  
 A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129;  
 A:Experimental source: splice form B  
 A:Note: Genbank entries HUMEPIS1 and HUMEPIS1B present only the amino-and carboxyl  
 R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burt  
 J. Biol. Chem. 265, 15286-15293, 1990  
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epith  
 A:Reference number: A35886; MUID:90368715  
 A:Accession: A35886  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-992,1033-1344 <GEN>  
 A:Cross-references: GB:J05581; NID:9188869; PIDN:AAA59876.1; PID:g188870  
 A:Note: Genbank entry HUMUCB includes one copy of the tandemly repeated sequence  
 R:Ian, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.  
 J. Biol. Chem. 265, 15294-15299, 1990  
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.  
 A:Reference number: A35887; MUID:90368716  
 A:Accession: A35887  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>  
 A:Cross-references: GB:J05581; NID:9189598; PIDN:AAA60019.1; PID:g189599  
 A:Note: Genbank entry HUMPAN2 contains four fewer copies of the tandemly repeated se  
 R:Wreschner, D.H.; Harevuoni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,  
 Eur. J. Biochem. 189, 463-473, 1990  
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen  
 A:Reference number: S10571; MUID:90276413  
 A:Accession: S10572  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <MBE>  
 A:Cross-references: EMBL:X52229; NID:g37053  
 R:Wreschner, D.H.  
 submitted to the EMBL Data Library, March 1990  
 A:Reference number: S40293  
 A:Accession: S40293  
 A:Molecule type: mRNA  
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <  
 A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054  
 R:Abu, M.; Siddiqui, J.; Kufe, D.  
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989  
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associat  
 A:Reference number: A36735; MUID:90088473  
 A:Accession: A36735  
 A:Molecule type: mRNA  
 A:Residues: 1142,'Q',144-162,'Q',164-168 <ABE>  
 A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543  
 R:Masuzawa, Y.; Miyachi, T.; Hamanoue, M.; Ando, S.; Yoshida, J.; Takao, S.; Shlmazu  
 J. Biochem. 112, 609-615, 1992  
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut ag  
 A:Reference number: JX0235; MUID:93123189  
 A:Accession: PX0066  
 A:Molecule type: mRNA  
 A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>  
 A:Experimental source: gastric carcinoma cell  
 R:Zilhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.  
 FEBS Lett. 356, 130-136, 1994  
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytolin  
 A:Reference number: S51026; MUID:95080414  
 A:Contents: annotation  
 A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region  
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 12  
 partial repeats. The repeat shown is defined by SmaI nuclease sites.  
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively  
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48

A:Accession: A26877  
 A:Molecule type: DNA  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: EMBL:M16164; NID:g172522; P1DN:AAA5014.1; PID:g172525  
 A:Accession: B26877  
 A:Molecule type: DNA  
 A:Residues: 762-1331 <YAA>  
 A:Cross-references: EMBL:M16165; NID:g172523; P1DN:AAA5015.1; PID:g172526  
 R.Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
 FBS Lett. 239, 179-184, 1988  
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
 A:Reference number: S27281; MUID:89031230  
 A:Accession: S27281  
 A:Molecule type: DNA  
 A:Residues: 1-31 <PAR>  
 A:Cross-references: EMBL:X13857; NID:94551; P1DN:CAA32069.1; PID:g4552  
 R.Lambrecht, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
 A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudoty  
 A:Accession: JG6123; MUID:96323237  
 A:Reference number: JG6123  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <LAM>  
 A:Cross-references: GB:U30626; NID:g1304386; P1DN:AAC49609.1; PID:g1304387  
 A:Accession: M1367 <LAM>  
 A:Cross-references: M1367 <LAM>  
 A:Gene: SGD:MUC1; STA2; MAU5; DEX2; SGD:S0001458  
 A:Cross-references: MIPS:YIK019C; SGD:S0001458  
 A:Map position: 9R  
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
 F:5-21/Domain: transmembrane #status predicted <TM>  
 F:1350-1366/Domain: transmembrane #status predicted <TM>

Query Match 15.2%; Score 762.5; DB 1; Length 1367;  
 Best Local Similarity 27.8%; Pred. No. 1.7e-25;  
 Matches 280; Conservative 115; Mismatches 441; Indels 171; Gaps 42;

```

QY 35 TTPTSTTQNHKVTSPKITTAKINP-----RSLPNSDTSKETSLLVKKETTYE 86
D 236 STSSSTSTSSSTSSSTSSSTTAPATPTTSCREKPTPTTSCREKPTPHNDHTPC 205
QY 87 TKETTTNNKOTSDGKERTTSKETSIAKETOSEKTSKADLPKTVLAKPKRAET--TKGRA 144
D 296 TKKTTTSK-TCT--KTTTPVPTPS--SSTTESSAPV-----PSSSTTESSAPV 344
QY 145 LTPKE-----PPTTPKPAATTPKPTTITKAP--TTPKPAATTTKSAPTTP 194
D 345 TSSSTSSAPVPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTPSSSTTE 404
QY 195 KEPA-----TTPKEPA-----TTPKEPA-----TTKSAPTTPKEPA 238
D 405 SSSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVPTPSS 464
QY 239 TTPKKAAPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTP 284
D 465 TSSSSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESS 524
QY 285 PTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 336
D 525 APAPVTSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPT 583
QY 337 EPAPTTTTSKAPPTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 390
D 584 PESSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPT 640
QY 391 KEPAATTTTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTP 446
D 641 PPSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVTS 697
QY 447 EELAPTTPEPTP--TTPKEPAATTPKPAATTPKPAATTPKPAATTPKPAATTP 493

```

```

D 698 STTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSST 757
QY 494 -----PKEPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 545
D 758 ESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSS 817
QY 546 TSDKAPATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 605
D 818 ITSSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSS 874
QY 606 TTKGP-----TSTSDKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 658
D 875 ITSSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSS 931
QY 659 STPTTKEPTTIHKSPDESPELSAEPKPALENSKPEGPVPTTPKPAATTPKPA 714
D 932 TTPSTTTITTTVCSTGNSAGETTSKSPKTVTT-----VPTTTTTSVTSTTIT 966
QY 715 ---TAKKTERDRLRTPET-TTAAKMKETATTEKTESITATTOVST--TTOD 768
D 987 VCSGTNSAGETTSKSPKTVTTTTPCSTSPSEASESTTSPPTTVTVSTTVTTT 1046
QY 769 TTPPK-----ITTLKTTTLA--PKYT--TKTITTEI--NKKPEPT-- 805
D 1047 STSKTPGCEITTTTTPVKNIPPTVLTTLIAPVSTVTVNPTTITTTVCSTGNSAGET 1106
QY 806 --AKPKRATNSKATTPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPKPK 863
D 1107 SCSPKTVTTTTPCSTGTEGTTTATTVTTTVCSTGNSAGETTVGTTTS 1166
QY 864 NPTSRIAEAMLOTTTRPNQ--TPNSKLYEVNPKSED--AGGEGET 905
D 1167 VPTTYV-----TTLAPSAVPTPATNAVPTTITTTTCSATNAGET 1207

```

Query Match 13.4%; Score 671; DB 2; Length 1151;  
 Best Local Similarity 25.9%; Pred. No. 1.1e-21;  
 Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

RESULT 8  
 T18535  
 high molecular mass nuclear antigen - chicken (fragment)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18535  
 J. Shimada, K.; Harata, M.; Mizuno, S.  
 J. Cell Sci. 110, 3031-3041, 1997  
 A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch  
 A:Reference number: 218955; MUID:9803440  
 A:Accession: T18535  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1151 <SHI>  
 A:Cross-references: EMBL:D84440; NID:d1177138; PID:d1025045; P1DN:BA424137.1

```

QY 14 TPKP-----PYVDEAGSLDNGDFKVTPTDSTYQHNKYST--PKITAKPINPPS 64
D 113 TTPPQAGPAGTPPQGAAGAKGCGTAPGSGTSGAAGKAAADVPKATTA--ATEAP 170
QY 65 LPNPDSTKEISLVYANKETVETKETTNNK--QSTDGKERTTSKETOSEKTSKADL 122
D 171 -----ASAASPTVPKATATATVTTAASOGAPAAATDAAVTAA--SQSAKATV 219
QY 123 APTSKVLAKPTPKAETTTGPAATTPKPAATTPKPAATTPKPAATTPKPAATTP 177
D 220 KPAATAVAKEKATVTAATAAKATAEAKPAPVTSPTTTPCSSAEAKPLTAASPTASKA--T 277
QY 178 TPKPAATTP-----TKSAPTTPKPAATTPKPAATTPKPAATTPKPAATTP 203
D 278 AEAKPVPTATSLMATKYTAEAKPAPSPSVKATTDFAAVATATAKAPDVKPAVAVCAEA 337
QY 204 EPAPTTTPKEPAATTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 250

```

```

Db 769 PPAPLSSPPAPQVYKSSPPVQVSSP-PPAPKSSPLPAVSSPPQVEKTSPPAPLSSPP 827
Qy 412 TTPKKTAP-----TTP-----KLLTPTTPKELAPTTPEKAPAP-----TTPPEELAPTTPEPT 458
Db 828 LAPKSSPPHVVVSSPPVYKSSPPAPVSSPPPLPKAPSPAHVSSPPPEVYKSPSP-PA 885
Qy 459 PTTPEEPAPTTTTPKKAAPNTPKK-PAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTT 517
Db 886 PTTVISP-PSEKSSPPPTPVSLPPIYKSSPPAMVSSP-PMTPKSSPPVAVSSPPPT 943
Qy 518 IKE---PA-----PTTPKK---PAKELAPTTKE---PTSTTSKPAPTTKGAPTTT 563
Db 944 VKSSPPAPVSSPPATPKSPSPAPVNLPPPEVYKSSPPPTVSSPPA---PKSSPPAP 1000
Qy 564 -KEPAPTTKEPAPTTPKATPTLKEPAPTTPKKAPAPLAPTTTKGTSTTSKAPAP 622
Db 1001 MSSPPPEVYKSSPPAPVSSPPVYKSSPPAPVSSPP---PPVYKSSPPAPVSSPPPV 1057
Qy 623 -TPKETAAPTTPEKAPAPTTTPKAPAPTTTPPTTSEVSTPTTKKEFTTHKSPDESTPL 681
Db 1058 KSPPPAPLSSPPPVYKSSPPAPVSS-SPPPVYKSSPPAPVSSPPPVYKSSPPAPV 1115
Qy 682 SAEPPTKALENSPEKGVPTTKTPATKPEMTTAKDKTTEDLRTTPT 731
Db 1116 SPPAP-----VKPSLP-PPAPVSSPPVTPAPPKKEOSLPPAP 1158

RESULT 6
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yotou, W.Y., St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 12187 <YOT>
A:Cross-references: EMBL:U048363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C:Gene: Naca
A:Map position: 10
A:Insertions: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activi
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 15.5%; Score 776; DB 2; Length 2187;
Best Local Similarity 27.7%; Pred. No. 7.2e-26;
Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps 49;

```

```

Db 1033 APKTAVSEIIPSQGVAVPLEISLPLEKTSKATPGEKSASSPKRSKTAGPKE---TPP 1089
Qy 258 KEPTTPPKKEPAPTTKEPAPTTKEPAPTAAPKAPTT---PKKAP-----TTPKEPAPTT 311
Db 1090 GGVAVAPPEISLPEKETPQONATPNEISLAASSOKSPKTSVPEKETPEGGVATAMPLEISAP 1149
Qy 312 TKEPSPTPK-----EPAPTTKSAPTTTKEPAPTTT-KSAPTTPK-----EPSPT 357
Db 1150 QKAPTAAPVKKOIPPEEDAVTLLASPLSPKASASTAAPKAPAPISGVAVSEISPS 1209
Qy 358 TKEPAPTTKEPAPTTPKKAPTT-PKE-----PAPTTKEPAP-----T 396
Db 1210 KTSKTAAPKENSATLPPKRSKPTAAPKETPATSSEGVAVAPSEISSPPTPASKGVPT 1269
Qy 397 TTKAPAPAPKEPAPTTTKEPAPTTPKLPTPEKLAAPTTPEK-PAPTTPEELAPTTPE 455
Db 1270 LTPKAPNALAE-SPASBKVKVTAAPETSTP-----SPQIKPKAGKESATPDS 1332
Qy 456 EPTPT-----PEEPAPTTPKAAAPNTPEKAPAPTTPEK-PAPTTKEPAPTTPEKTA 506
Db 1323 KATPATAAPKETASSEGVAVPLEIPSPKAPATAAPKETPAPS--PEGATTAPOIP 1380
Qy 507 PTPKGTAPTTLKEPAPTT-----KPAKELAPTTTKEPS-- 544
Db 1381 PSPKSGSKAGSKE-TPTTPSEGVTAAPLEIPSSKTSKASKEPLVTPSSKLSOT 1439
Qy 545 -----TTSOKAPATTTKGTAPTTKEPAPTTPEK-PAPTTKGTAPTTLKEPAPTTPKKA 599
Db 1440 VGPKEISLEGATAVPLEIPSPKAKPVYDQVLTSPK-DAPTTLAE-SPSSPK-A 1496
Qy 600 KEELAPTTKGTSTTSKAPAPTTKETAAPTTPEKAPAPTTPKKAPPTTPTP-----PPT 654
Db 1497 PETAAPPSR-VTVYPEKPA-TPKASGTSKVPVPEVPEVAVSSHEPVTPEAPV 1554
Qy 655 TSEVSTPTTK-----EPTTHKSPDESTPELSAEPKALENSPEK-PGVPTTKTPAA 707
Db 1555 KNSSHKTSKTELEKAPATLPPSPKSPKIPSSKABRT--SAKPEPASPSTK-PVT 1611
Qy 708 TKEMTTTKDKTTEDLRTTPETTAAPKMTKETATTEKTESKITATTTQVSTTTQ 767
Db 1612 T-SLQOTAPPSLOKAPSTTIKRENLAAPV-----LPVSSKAPAPARASASLPTAAP 1665
Qy 768 DTPPKITTLKTTTAAKVTTKITTTTEINKPEETAKPKDRATNSKATTPKPKTK 827
Db 1666 QTPAKBATTIPSCKKAATETIETSTASLDEAPKETSE---TSVSVLMSSP----- 1716
Qy 828 APKKPSTKKPKMPVVRKKTTPTRKMTSTMPLENPISRLAEAMIQTTTPNQTPNSK 887
Db 1717 -PKASSSKASTLP-----ATLPSLKEASVLS-----PTATSSGK 1752
Qy 888 LVEVNPKESEDAGAGEETP 906
Db 1753 DSHSPVS-DACSTGTTT 1770

RESULT 7
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01
C:Species: Saccharomyces cerevisiae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; J6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Molecule type: DNA
A:Residues: 11367 <ROW>
A:Cross-references: GB:247047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN00009;
R:Yamashita, T.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A:Reference number: A91831; MUID:87194600

```

## RESULT 4

T16251

hypothetical protein F35A5.1 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000

C:Accession: T16251

R:Leinbach, D.

submitted to the EMBL Data Library, January 1996

A:Description: The sequence of *C. elegans* cosmid F35A5.

A:Reference number: 218485

A:Accession: T16251

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1274 &lt;LEI&gt;

A:Cross-references: EMBL:U46675; NID:q116613; PID:q116621; PIDN:AA82641.1; GSPDB:GNOC

A:Experimental source: strain Bristol N2; clone F35A5

C:Genetics:

A:Gene: CESP:F35A5.1

A:Map position: X

A:Introns: 1272/2

## Query Match

Best Local Similarity 16.0%; Score 802; DB 2; Length 1274;

Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```

OY 6 KNRTRK-KP-----TP-----KPVVDAGSLDNGD---FKVTPPTSTTQHNVKSTS 50
DB 274 KNPTRKMKRPWEDETPVEEVEKEPVPKAPVKKKDDPAKAKANDPSPKAAKVPK 333
OY 51 PKITTAAPKPI-NRPSLPRNSDSKETSILVNKETIVERKETTITTKQSTOCKETTSAK 109
DB 334 SPVVEPTPVKNVKKKKPPWEVDDEPAEVEKKPSAEKPTVLRKKPEPSTTPSSDPS 393
OY 110 ETQSTSEKTSANDLAPTSVLAKPPPKAE---TTTKGPA-----LTTTP 148
DB 334 PKKAPAVKPRDSSPKKATPIQADPKAQEVPPTPVKNVKKKKPPWEVDDEPVEVKQP 453
OY 149 KEPTPTTP---KEPASITPKPEP---TP-TTISAPITPKPEAPTTTKSAPTPKPEAP 199
DB 454 EAPAKKTIVLRKKEPAADTKPATSKTPEPEKKDPVKPRDPSKKVAAKPDQAQAPA- 512
OY 200 TTTKCPA-----PTTKPEAPTTTKEPAPT-----TTKSAPITP 233
DB 513 TPVKNPVKMKRPWEDETPADVDVSKPTDAKKTPSLAKKDDAPAKESLKPADTKAAR 572
OY 234 KEP-----APTTPKKRAPTTTKE---PAPTTPKEPTPTTTPKEPAPTTPKEPAPTTP 283
DB 573 RDPSEKKVAPTAPEKKTTPVLAKKEPAGPADSKTEPEKSKRDPSPKAVPAKVPKTEV 632
OY 284 APTAKKRPAP-----TTPKEPAPTTPKEPAPTTPKEP---SPTTKEEAPTTTKSAPT 333
DB 633 APAVVKKEPISPKPDAPAKAEKPNSEVVP-PTPVKNPVKMKRPWEDDDAPKAVSLPE 691
OY 334 TTKPEAPTTTSAPTTP-KEPSPTTKEPAPTTPK--EPAPTTPKKAP--TTPKEPAPT 388
DB 692 PEKKT--TVLAKKATPKDSEAAADPVSGPSSKDKLAKKAPVKRDPSPKAVPIKAPK 750
OY 389 TPKEPAPTTPKAPTA--PKEPAPTTPKEPAPTTPKKTLPPTP---EKLAPTTPEKPA 442
DB 751 T--EVPAAVVKKEPVAKSDPSKKK--AEPSN--VVPPTPVKNPVKMKRPWEDDDA 805
OY 443 PTTPEELAPTTPEEP-----TPTTPEEPAPTTPKAAAPNTPKEPAPTTPKEPAPTTP 496
DB 806 PAEAVNVPBEKKTTPVLAKKTVPKPRDPSPKKAVAPSKTKTDAPVSVKKEPVEKPK 865
OY 497 PAPTTPKETAPTPKGTAPTTLKEPAPTTPKKAAPKELAT--TTKEPTSTTSKAPPTTP 555
DB 866 PSRKKAAKPNSEVVP---PTPVKNPVKMK--KRPWEDDEDETEVEKKPSE--PEKKTTPVLA 918
OY 556 KGTAAPTTPKAPAPTTPKEPAPTTPKGTAPTTLKEPAPT-----TPKKRPAP-----KEI 603
DB 919 K-KEPEKPD-APKVAKPRDPSKKAVPE--KEPAKVAKPRDPSKKAIPIRANTQA 974

```

```

OY 604 APTTKGPTS-----TTSDKP-----APTTPKET-----APTTPKEPAPTTPKAP 646
DB 975 PPTPVKNPVKMKRPWEDDEPAEPVSAPEPEKKTTPVLAKKAPKAPKRP---SRKAAV 1031
OY 647 TPETPPTTSEVSTPTTTPKETTTHKSP---DESTPLSA-EP---TPKALENSPKKEG 698
DB 1032 AAK-PDPIPEV-PPTPVKNPVKMKRPWEDDESEVSAPEPEKKTTPVLAKKAPKPA 1089
OY 699 V-----PTTKPAAT-----KPEMTTAKDKTTEDLRTPEPT--TAPK 737
DB 1090 TKPDSEAAADPVSGPTSDPKLSKKAPVEKKTPTDPRDGLKPSPAKPEKAPAPK 1149
OY 738 MTKETATITKTTESKITATTGVTSTTODTPEFKITTLTKTTLAPKVTTLKKTITTE 797
DB 1150 KKKPVWDDPDPEPADFTVPAPSKRPDDEPADPLG-----GPKTDPK----- 1193
OY 798 IMNKPEETAKPKDRATNSKATTPKPKPKPTSTKPKTPPVKPKK----- 848
DB 1194 -LNKKAPAEKPTK-----PKPEVSKPEPKPTPEPKR-APKMKRPWEDDDEPE 1243
OY 849 ---TTPPRKMTSTMP 861
DB 1244 ADFTPAPAKKPDTEDP 1259

```

## RESULT 5

S49915

extensin-like protein - maize

C:Species: *Zea mays* (maize)

C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999

C:Accession: S49915

R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.

submitted to the EMBL Data Library, June 1994

A:Description: Pex genes: pollen-specific genes with extensin-like domains.

A:Reference number: S49915

A:Accession: S49915

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1188 &lt;RUB&gt;

A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

## Query Match

Best Local Similarity 15.6%; Score 784; DB 2; Length 1188;

Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```

OY 13 PTPKPVVDAGSLDNGDKVTPPTSTTQHNVKSTSPKITTAAPINRPSL---PPN 68
DB 460 PTPHSPAPD-----DYVPTTPVPKSPSPATSPSPQVQPPAASPPPSLVLSPPQ 510
OY 69 SDTSKETSILVNKETIVERKETTITTKQSTOCKETTSAKETQSIEKTSAKDLAPTSKV 128
DB 511 APVG-----SPPPVKTTSPAPIG---SPSPPPVSIV 541
OY 129 -----LAKTPKAETTTKGPALTTPKEPTTPTPKEPASTTPKEPTTPITIKSAPTTKEPA 183
DB 542 SPPPVKSPPPAPVAGSPPPEKSPPPAPVAVSPPPVKSP--PPPTLVASPPPVKSP 599
OY 184 PTTTKSAPITP--KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTP--KEPAPT 240
DB 600 PAPVAVSPPPVKSPPPPTPVASPPPAVAVSSPPPMKSPPPPTVSSPPPEKSPPPPP 659
OY 241 PKKAPAPTTPKEPAPTTP-----PK-----PTPTTPKEPAPTTPKEPAPTTPKEPAPTTP 291
DB 660 PAKSTPPEPEYPTPTSVKSSPPPEKSLPPPTLISPPOKEPTPESTPSKP--PSSPEK 718
OY 292 APTTPKEPAPTTPKEPAPTTPKESPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPK 351
DB 719 SP--PKEPVSSPQTP-----KSSPPAVSSPPPTPVVSSPALAVSSPPVSKSP----- 768
OY 352 EESPPTTKEPAPTTPKEPAPTTPKKAAPTTTKEPAPTTPKKEPAPTTPKKAAPTTTKEPAP 411

```

Db 865 SDEPTSPDEPTSP--DEPTSP--DEPTSP--SDEPTSPDEPTSPDEPTSPDEPTSP 919  
QY 302 TTPKPEPA--PTTKPESP--TKPEAPATTTKS--APTTPKEAPATTTKSAPTTPKESPTT 357  
Db 920 SDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 968  
QY 358 TKEPAPATTKEAPATTPKKPAATTPKEAPAT--TKPEAPATTTKKPAATTPKEAPATTPK 416  
Db 969 SDEPTSP--DEPTSP--DEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1019  
QY 417 TAPTPPKLPTTPPEKLAATTPKEAPATTPEELAPTTPPEPTT--TPKEAPATTPKAAAP 475  
Db 1020 -----EPTSPDE--PTSPDEPTSPD-----EPTSPDEPTSPDEPTSPDEPTSPDE 1064  
QY 476 NTPEKAPATTPKPEPA--PTTKPEAPAT--PKPEAPATTPKGATATTPKEAPATTPKAPK 532  
Db 1065 TPSEDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1122  
QY 533 ELAPTTPKEPT--STSPDKAPATTPKGATATTPKEAPATTPKGATATTPKEAPATTPK 591  
Db 1123 D-EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1175  
QY 592 PT--TPKPAKELAPATTTKAPTSTSPKAPAT--TPKEAPATTPKKAPAT--TPKAPAT 646  
Db 1176 PSETPPEEP-----IPDTPSDEPTSPDEPTSPDEPTSPD--EPTSPDEPTSPSETPPEE 1229  
QY 647 TPETPTTSEVSTPTTKPEPTTHKSPDESSTPELSAAPTTPKALENSPKKEGVTTKTPA 706  
Db 1230 DTPEDEPTSPD--EPTSPDEPT-----PSDEPTP--SDEPTP--SETPPEEP--IPTDTPSD 1277  
QY 707 ATKPEMTTAKDKTTERDOLTPPTTTAAKMKETATTEKTESKITATTOVSTTT 766  
Db 1278 EPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1337  
QY 767 QDTPPTTKTTKTTLAPKVTTKTTTTEIMKPEETAKPKDRATNSKATTPKPO--KP 825  
Db 1338 DEPTSPDEPT-----PSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 1372  
QY 826 TKAPKKTSTKPKTPMPVRKKKTPTPPKMTS-----TMELNPTSKIA 870  
Db 1373 SEPTEPEPTTTP-----TPTSTPTSPSGSGSGSGSGGGGGGATVPSTPTPTPTSP 1424  
QY 871 EAMLOTTTRPNOTRPSKIVNPKSEDAAGAGETPHMLLRPHYEMPEVTDMOYL 927  
Db 1425 -----KPTSTAPATPEIE--EPTSPDVGAIGGHRAYLKG-----PDGSTRP 1465

RESULT 3  
T31108  
cyst germination specific acidic repeat protein precursor - Phytophthora infestans  
C:Species: Phytophthora infestans (potato late blight agent)  
C:Date: 22-Oct-1999 #sequence:rev1sion 22-Oct-1999 #text:change 22-Oct-1999  
R:Accession: T31108  
R:Goernhardt, B.  
Submitted to the EMBL Data Library, April 1998  
A:Reference number: Z20986  
A:Accession: T31108  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1489 <GOE>  
A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AA072308.1  
C:Genetics:  
A:Gene: car90

Query Match 17.0%; Score 853; DB 2; Length 1489;  
Best Local Similarity 32.3%; Pred. No. 2.9e-29;  
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TTKKPTPKPPVVDGSGDNGDFKVTTPDTSTT-----QHNKYSTSKITATKAPINR 62  
Db 338 TPAPTEKPYDVETTYVTEESTYAPTCKSETNAPTERMHYAHLEKPCDTEVTWYAPTEET 397

QY 63 PSLP-----PNSDT-----SKETSLVNKEETTVETKETT-----TNKQSTSD 100  
Db 398 TPAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 457  
QY 101 GKEKTTSAKETOSIKTSKADLAPTSKYLAKPEATTTKQALTYTPKEPTTPPKKPA 160  
Db 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514  
QY 161 STTPKEPTPTTKSAPTTPKEAPAT--TKSAPTTPKEAPAT--TKPEAPATTPKKAPATT 217  
Db 515 EETPYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 573  
QY 218 TKEPAPATTKSAPT-----TPKEAPATTPKKAPATTPKKAPATTPKKAPATTPK 267  
Db 574 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 632  
QY 268 P-----AP-----TKPEAPATTPKK-----APAPKPKAPATTPKKAPATT 304  
Db 633 PTEETTYASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 692  
QY 305 KEAPAT--TTKESPTTPKEP-----APT--TKKSAPT-----TTKEP-----A 339  
Db 693 TTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 752  
QY 340 PT--TKKSAPT--TPKESPTTPKEP-----APTTPKEAPATTPKKAPATTPKKAPAT 388  
Db 753 PTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 812  
QY 389 TPKEAPATTTKKAPAPKAPATTPKEATTPPKLPTTPPKLAPTTPKEAPATTPEE 448  
Db 813 TPYEPTTEETTYAPTEETTYAPTEET--TYTPTEETTYAPTEETTYAPTEETTYAPTEET 870  
QY 449 LAPAT--TPPEPTTPTEEP-----APT-----TPKAAPNTPKKAPATTPKKAP 489  
Db 871 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 930  
QY 490 -----APTTPKEAPATTPKKAPATTTPKGTAPT--TLKEAPATTPKKAPAP--ELAPT--TT 539  
Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 990  
QY 540 KEPTSTSDKAPATTTPGATAPTTPKEAPATTTPKAP-----APTTPKGATP 585  
Db 991 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1050  
QY 586 -----TLKEAPATTPKK-----APKE-----LAPT--TTKGTSTTSKAPATT 624  
Db 1051 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110  
QY 625 KETAPATTPKKAPATTTPKAPATTTPETTPPTTSEVSTPTTKKEPT--TIHKSPEDESPELS 682  
Db 1111 YAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1169  
QY 683 AEPKPKA--LENSPKPEPG-----VPTTPKPAATKPEMTTAKDKTTERDOLKTP---ETT 732  
Db 1170 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1229  
QY 733 TAAPKMKETATTEKTESKITATTOVSTTTODTT-----PKKIT--LKT 779  
Db 1230 TYAP-----TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1284  
QY 760 TTLAKVTTKTKLITTEIMKPEETAKPKDRAT--NSKATTPK--OKTPAKPKPT--- 833  
Db 1285 TTYAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1344  
QY 834 -----STKPP--KTMPVRKKKTPTPPKMTSMTLNPSTIAEAMLOTTTRP--N 881  
Db 1345 PAESTSTYSTSKPCNTPEPTDEPTDPE--PSDEPTDEPTDPLDPEDEPTPCDN 1402  
QY 882 QTPNSKIVNPKSDDAG 899  
Db 1403 QGINGIGVENKVRVNNAG 1420

J. Clin. Invest. 87, 77-82, 1991  
 A:Title: Human bronchus and intestine express the same mucin gene.  
 A:Reference number: A61257, MUID:91086481  
 A:Accession: A61257  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1,1925-1948, 'TTS', 1952-1954 <JAN>  
 A:Experimental source: bronchus  
 R:Xu, G.; Huan, L.; Khatiri, I.; Sajjan, U.S.; McCoil, D.; Wang, D.; Jones, C.; Forstner, Blochem. Biophys. Res. Commun. 183, 821-828, 1992  
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus  
 A:Reference number: P00328, MUID:92198477  
 A:Accession: P00328  
 A:Molecule type: mRNA  
 A:Residues: 2328-2468 <XUG>  
 A:Cross-references: GB:M85523  
 A:Experimental source: small intestine  
 A:Accession: P00329  
 A:Molecule type: protein  
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>  
 A:Gene: GDB:MUC2  
 A:Cross-references: GDB:120203; OMIM:158370  
 A:Map position: 11p15.5-11p15.5  
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von F.2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 21.2%; Score 1062.5; DB 2; Length 3020;  
 Best Local Similarity 29.2%; Pred. No. 8e-38;  
 Matches 342; Conservative 74; Mismatches 442; Indels 315; Gaps 39;

```

QY 32 EVVTPEDSTQH---NKVSTSPKTTAKPINRPSLPSNDSKETSILYKNETVETK 88
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1240 WEICPNEGVEKHNICSTTRPSLTFTFTTLPTTP-----TSFTTTTTTTPTS 1291
QY 89 ET--TTTNK-----QTSNDGKEKTS--AKETOSIEKTSAD----- 121
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1292 STVLSTTKLCLCWSMDWINEHPSGSDGDREPDGCGAEDIECNSVDPHLSLOH 1351
QY 122 -----LAPSKVLAKPPKAE 137
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1352 GOKVOCDSVGFICKNEDQNGFGLCYDKIRVNCQWMDKITTSPPTTTPSPPT 1411
QY 138 TTTGPAITTEKEPPTTPKPASTTKEPPTTIKSAF--TPKEPATTTKSAFTTPE 196
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1412 TTTTLEPPTTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1471
QY 197 PAPTTEKEPATTTEKEPATTTEKEPATTTKSAF--TPKEPAPTTPKKAPATPEPAP 255
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1472 P-PTTTPSPPTTSPPT--TTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPT 1529
QY 256 TPKEPPTTPKEPAPTTEKEPATTTPKAPPAKPAATTPKEPAPTTPKEPAPT 315
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1530 TTPPEPTTTPSPPTTTPITP--PTSTTTLPTTTPSPPTTTPPTTTPPTTTPSP 1588
QY 316 SPPTKEPAPTTPKSAF--TTTKEPAPTTPKSAF--TPKESPTTTPKEPAPTTP 373
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1589 PTITTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1648
QY 374 PKKAPATTEKEPAPTTPKEPATTTPKKAPATTEKEPAPTTPKEPAPTTPKE 433
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1649 TTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1700
QY 434 APPTPEKAPATTEBELAPTEEPPTTPTEEPATTPKKAAPTTPKEPAPTTPKE 493
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1701 PTTPSSPTTTPPTTTPPTTTP--SPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1759
QY 494 PKEPAPTTP-----KETAPTTP-----KG----- 512
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1760 PLPPSTTPPTTTPPTTTPPTTTPCVPCLCNWTGMLDSGKPRNHRGDTLIGDVC 1819

```

```

QY 513 ----- 512
Db 1820 NISCRATWPDVPIQLGQTVCDVGLICKNEDQKRGVITPAFCLNFINVQCECV 1879
QY 513 TAPPTLKEPAPTTPKPAKELAPPT--TKEPSTTSDDK---APT--TPKGAFTTPKE 565
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1880 TQPTTW-----TTTTENPPTPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1936
QY 566 PAPTTEKEPAPTTPKGAFTTLKEPAPTTPKPAKELAPPTTTPGPTST--TSKPAFTTP 624
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1937 PTPGTQTPPT--TTTPTTTTTTTPPTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1995
QY 625 KEPAFTTPKEPAPTTPPKP-----APTPEPTTPPTSEVSTPTTPK-----EPTTHK 672
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1996 ITTTTTPVATPTPTGQTPTPTWIPISITTTTTPPTPTTTPGTPPTTHSTAPAEITSNP 2055
QY 673 SPDSSTPLSAEPKPALENPKREPGVPTTKPATTPKMTTAKDKTTRDLRTTPPT 732
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2056 PPESTTPQTSSTSPLESTTLSTLTPALMTSTAPSTTPAPPTTSGHTLSPPST 2115
QY 733 TPAKMTKEPATTTTEKTESKITATTTQVSTTTQDTTPPKITT--LKTTLAPKVTTP 789
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2116 TTSPTGTPTRGTTP--GSSAPPTSTVQTTTTSAMPTTPPLSTPILRTTGLRPTSSV 2173
QY 790 KKTITTEINKKPE-----TAKPKDRATNSKATTP 820
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2174 LICCVLDITYAEGEEVYNGTYGDTCYFVNCSLSTLEFYNMSCSPSTPTPPSK--STP 2232
QY 821 KPOKTPAKPKPTSTKPKTMPRAVRKPTTPTRKMTSTPELNPSTRIEAMLOTTTP 880
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2233 TPKSPSTSPKPPNGTPPCPDPPR-----QENETWMLCDDCPW--ATCKY 2278
QY 881 NQPNKSLVEVP-----KSEDAAG-----ABGETPHML----- 910
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2279 NMTVEIKVACEPPMPETCSNGLOPVAVEBDPCWHMECDXCCTGMD--PHVTFDGLY 2337
QY 911 -----RPHVMEPTPDMOYLPRVNOGILIN 937
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2338 YSYOGNCTIYVLEISPSVD-----NFGYIID 2364
```

RESULT 2  
 T18262  
 S-layer protein - Clostridium thermocellum  
 C:Species: Clostridium thermocellum  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18262  
 R:Fujino, T.; Beguin, P.; Aubert, J.P.  
 J. Bacteriol. 175, 1891-1899, 1993  
 A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose.  
 A:Reference number: Z18847; MUID:93209931  
 A:Accession: T18262  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1664 <FCU>  
 A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 18.8%; Score 942; DB 2; Length 1664;  
 Best Local Similarity 33.0%; Pred. NO. 5.6e-33;  
 Matches 276; Conservative 83; Mismatches 312; Indels 166; Gaps 43;

```

QY 128 VLAKPTP--KAETTTGPAITTEKEPPTTPKPASTTKEPPTTIKSAFTTPEPAPT 186
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 758 VVIOAPAIKASDEPIPDTPSDPTPS-----DEPTPS---DEPTPSDEPTPSD 804
QY 187 TKSAPTTEKEPAPTTPKPAFTTKEPA-----PTTKEPAPTTPKSAFTTPKEPAPTTPK 242
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 805 EPTPSETPEEPIDPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSEEPIDPT 864
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 243 KPAFTTEKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 301
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:30 : Search time 82.85 seconds  
(without alignments)  
865.181 Million cell updates/sec

Title: US-09-556-246-1\_COPY\_200\_1140  
Perfect score: 5011  
Sequence: 1 VKDNKKNRKTKKPPV.....DMDYLPRVNOGIINPMLS 941

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1062.5	21.2	3020	A43932	mucin 2 precursor, S-layer protein -
2	942	18.8	1664	T18262	cyst germination s
3	853	17.0	1489	T31108	hypothetical prote
4	802	16.0	1274	T16251	extensin-like prot
5	784	15.6	1188	S49915	nascant polypeptid
6	776	15.5	2187	T30826	glucan 1,4-alpha-g
7	762.5	15.2	1367	S48478	high molecular mas
8	671	13.4	1151	T18535	mucin 1 precursor,
9	666.5	13.3	1344	A53175	hypothetical prote
10	659.5	13.2	1229	T25697	extensin class 1 p
11	633	12.6	489	T11622	hypothetical prote
12	632	12.6	761	C84672	hypothetical prote
13	632	12.6	7962	T38346	elastic titin - hu
14	629	12.6	3507	T34513	hypothetical prote
15	626.5	12.5	990	T15618	nucleolar phosphop
16	622.5	12.4	839	F75518	hypothetical prote
17	622.5	12.4	6642	T29757	protein UNC-69 - C
18	620.5	12.4	971	T19431	hypothetical prote
19	617	12.3	3570	T45025	mucin MUC5B, trach
20	607.5	12.1	801	T29018	hypothetical prote
21	607.5	12.1	924	S27923	gene LF3 protein -
22	605	12.1	379	S50125	larval glue protei
23	559.5	11.2	350	S22456	hydroxyproline-ric
24	556.5	11.1	856	T16543	hypothetical prote
25	555.5	11.1	875	S23760	polyphenolic adhes
26	555.5	11.1	1630	A53577	ascites stialoglyco
27	551	11.0	620	S06733	hydroxyproline-ric
28	549	11.0	2332	T34434	hypothetical prote
29	545	10.9	873	A47283	calpocin - fruit

30	542	10.8	369	2	S20500	hydroxyproline-ric
31	538.5	10.7	416	2	J00465	extensin precursor
32	533	10.6	1087	1	QFMSH	neurofilament trip
33	532	10.6	756	2	T27642	hypothetical prote
34	530.5	10.6	1162	2	JH0557	exo-alpha-stalidas
35	530	10.6	865	2	A47282	calcium-binding pr
36	522	10.4	328	2	J00985	hydroxyproline-ric
37	518	10.3	813	2	S70795	vsaA protein precu
38	518	10.3	1072	1	A37221	neurofilament trip
39	512	10.2	866	2	T45462	membrane glycoprot
40	509.5	10.2	662	2	A45155	mucin FIM-C.1 - Af
41	509.5	10.2	867	2	T45463	membrane glycoprot
42	509	10.2	1459	2	T32271	hypothetical prote
43	505	10.1	700	2	A54641	interspersed repea
44	504.5	10.1	606	2	A43427	neurofilament trip
45	503.5	10.0	1118	2	A48292	mucin, tracheobron

## ALIGNMENTS

RESULT 1  
A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C:Date: 10-Mar-1993 #sequence, revision 12-Apr-1996 #text, change 05-Nov-1999  
C:Accession: A49963; A45106; B45106; A43932; B35532; A61257; P00328; P00329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A49963; MOID:94132002  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GU1>  
A:Cross-references: GB:121998  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MOID:93016075  
A:Accession: A45106  
A:Molecule type: mRNA  
A:Status: not compared with conceptual translation  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A:Molecule type: mRNA  
A:Status: not compared with conceptual translation  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398  
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, J. Clin. Invest. 88, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MOID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIP:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eva  
A:Reference number: A33532; MOID:89197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874  
A:Experimental source: intestine  
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.



---



RESULT 14  
US-09-083-116-2  
; Sequence 2, Application US/09083116  
; Patent No. 6203795  
; GENERAL INFORMATION:  
; APPLICANT: CHAMBON, Pierre  
; APPLICANT: KIENEY, Marie-Paule  
; APPLICANT: LATHE, Richard  
; APPLICANT: HAREVEUNT, Mara  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
; TITILE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:

PRIOR APPLICATION DATA: 08/479,537  
 APPLICATION NUMBER: 08/479,537  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/FR91/00835  
 FILING DATE: 23-OCT-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/039,320  
 FILING DATE: 04-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/403,576  
 FILING DATE: 14-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Teskin, Robin L.  
 REGISTRATION NUMBER: 35,030  
 REISSUE/DOCKET NUMBER: 017753-025  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2035 amino acids  
 type: amino acid

```

1 STRANDEDNESS: single
2 TOPOLOGY: linear
3 MOLECULE TYPE: peptide
4 FEATURE:
5 NAME/KEY: Peptide
6 LOCATION: 128..1899
7 /note= "The amino acids spanning
8 OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
9 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
10 OTHER INFORMATION: repeats varies from 1 to 40."
11 FEATURE:
12 NAME/KEY: Peptide
13 LOCATION: 134
14 /note= "Amino acid 134 is X1 - Xaa
15 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
16 OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
17 FEATURE:
18 NAME/KEY: Peptide
19 LOCATION: 144
20 /note= "Amino acid 144 is Y - Xaa
21 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
22 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
23 FEATURE:
24 NAME/KEY: Peptide
25 LOCATION: 147
26 /note= "Amino acid 147 is X2 - Xaa
27 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
28 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
29 FEATURE:
30 NAME/KEY: Peptide
31 LOCATION: 1..21
32 /note= "Amino acids 1 to 21 are a
33 OTHER INFORMATION: 21 amino acid precursor sequence."
34 US-09-083-116-2

```

Query Match	6.2%	Score 424.5	DB 4	Length 2035
Best Local Similarity	23.7%	Pred. No.5.3e-20		
Matches 255	Conservative 69	Mismatches 511	Indels 241	Gaps 53
QY	140	TSLVNKEETVEKETETTNKOTSDGKEKTSAKETOSIEKTSAKD-LAPTSKVLAKPT	198	
Db	16	TVLVV-----VTSGHASTPGEKEKTSIAQRSSVPSSTEKMAVSMTSVLSGIS	65	
QY	199	P-KAETTTKGP--ALTTPKP-----TPPTTKPEASTTKERTPTTKSAP	241	
Db	66	PGSGSSTTGODVTLAPATERPAGSAAATMGODVTVTRALGSTTPRANHVT---SAP	122	
QY	242	TTTKPEARPTT-----KSAPTTTKP-----APTTKEPARPTTKPEARPTT	282	
Db	123	--DNKPAAGSTARXAHNGVTSAPDXKRGSGSTARXAHNGVTSAPDXKRGSGSTARXAHNGVTS	180	
QY	283	TKE--PAVTT-----KSAPTTKEPARPTTK-----KPAPTTKKEPARPTTKERTPT	329	
Db	181	APDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVTS	240	
QY	330	PKE--PAVTTKP-----APTTKEPARPTAPK-----KPAPTTKKEPARPTTKPEARPTT	376	
Db	241	APDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVTS	300	
QY	377	TKESPTTKPKPARPTT--TKSAPTTTKPEARPTT-----TKSAPTTTKERSPTTK---E	425	
Db	301	APDXRRP--XPGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVT	359	
QY	426	PAPPTTKPEARPTTK-----KPAPTTKKEPARPTTK-----EPAPTTTKKRAPATAKREPARPT	477	
Db	360	SAPDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVT	419	
QY	478	TPKETA-----TPPKLLPTTPEKLAETTPKRAPPTTPEELAPPTPEEPPTTPEEBAPPT	533	
Db	420	SAPDXKRXGSGSTARXAHNGVTS---APDXKRXGSGSTARXAHNGVTSAPDXKRXGSGSTARXAHNGVT	471	
QY	534	TPKA-----AAPNTPKKEPARPTTKPEARPTTKPEARPTTKETAPPTPKGT--APTTLKEBPAP	588	

DB 958 VTSAPDXRXPGSTAPXAHGVTSA-----PDXRXPGSTAPXAHGVTSA PD 1003

## RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREUYEN, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 01753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note= "The amino acids spanning

OTHER INFORMATION: /note= 1899 constitute a repeated region wherein the repeat

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

OTHER INFORMATION: Xaa Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa

LOCATION: 144  
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,  
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 147  
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,  
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..21  
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
OTHER INFORMATION: 21 amino acid precursor sequence."  
US-08-479-537A-2

## Query Match

Best Local Similarity 23.7%; Score 424.5; DB 2; Length 2035;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

DB 140 TSLTVNKETVETKETTITTKOTSTOCKETTSKAKETQSIKTSKAD-LAPTSKVLAKPT 198  
16 TVLTV-----VYSGHASTPGEKETSATQSSVSPSTKNAVTSLSHS 65  
QY 199 P-KAETTTGP--ALTTPKEP-----TPTPKEPASTTPKEPPTTIKSA 241  
DB 66 PGSSSTTQODYTLAPATPASPASAAATWQODVTSVTPPALGSTTPPAHDVT---SAP 122  
QY 242 TTPKEPAPTTT-----KSAPTTPKEP-----APTTPKEPAPTTPKEPAPT 282  
DB 123 --DNKPAPGSTAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 180  
QY 283 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KAPTTTPKEPAPTTPKEPAPT 329  
DB 181 APDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 240  
QY 330 PKE--PAPTTKEP-----APTTPKEPAPTAPK-----KAPTTTPKEPAPTTPKEPAPT 376  
DB 241 APDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 300  
QY 377 TKEPSPPTTPKEPAPT--TKSAPTTTPKEPAPT-----TKSAPTTTPKEPAPTTPK-----E 425  
DB 301 APDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 359  
QY 426 PAPTTTPKEPAPTTPK-----KPAPTTPKEPAPTTPK-----BPAPTTPKAPTAKEPAPT 477  
DB 360 SAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 419  
QY 478 TPKEPTAP---TPPKLITPTTPKEKLAPTTTPKEPAPTTPPELAPTTTPPEPTTPTEEPAPT 533  
DB 420 SAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 471  
QY 534 TPKA---AAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 588  
DB 472 APXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 530  
QY 589 TTP-----KKPAKELAPTTTPKEPTSTSDKPAP--TTPKGAPTTTPKEPAPTTP 636  
DB 531 TAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 585  
QY 637 KEPAPTTPKG---TAPTTTLKEPAPTTP-----KKPAKELAPTTTPKGTSTSD 682  
DB 586 PXPGSTAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 644  
QY 683 KPAP--TTPK---ETAPTTTPKEPAPTTP-----KKPAPTTPPTTPPTTPSEVSTP 726  
DB 645 RXPGSTAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 702  
QY 727 TTPKEPPT-----IKSPDESTPELSAEPPTPKALENSPKEPPTTKTPAA----- 772  
DB 703 DXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVTAPDXRXPXGTAAPXAHGVT 762

[illegible]

APPLICATION NUMBER: WO PCT/FR91/000835  
 FILING DATE: 23-OCT-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/039,320  
 FILING DATE: 04-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/403,576  
 FILING DATE: 14-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Teskin, Robin L.  
 REGISTRATION NUMBER: 35,030  
 REFERENCE/DOCKET NUMBER: 017753-025  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1867 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 128..1727  
 OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat  
 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such  
 OTHER INFORMATION: repeats varies from 1 to 40."  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 134  
 OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa  
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 OTHER INFORMATION:  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 144  
 OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa  
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA  
 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."  
 OTHER INFORMATION:  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 147  
 OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa  
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA  
 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."  
 OTHER INFORMATION:  
 FEATURE:  
 NAME/KEY: Peptide  
 LOCATION: 1..21  
 OTHER INFORMATION: /note= "Amino acids 1 to 21 are a  
 OTHER INFORMATION: 21 amino acid precursor sequence."  
 US-08-479-537A-5

Query Match 6.2%; Score 424.5; DB 2; Length 1867;  
 Best Local Similarity 23.7%; Pred No. 4.8e-20;  
 Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 140 TSLTVKKEVETKETTNNKQSTDKKETSASKEQSIKTSKAD-LAPTSKYLAKPT 198  
 | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 16 TVLTV-----VTGSGHASTPGGKETSATGRSSVPSSTENKNAVSMSTVLSHS 65  
 QY 199 P-KAETTTKGP-ALTTPKP-----TPTPKKEASTTPKEPTTTIKSAP 241  
 | : | | : | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 66 PGGSGSTTGQDVTLPATEPAGSAAATWQDVTSPVTRPALGKSTTPRAHDVT---SAP 122  
 QY 242 TTRKEPAPTTT-----KSAPTTKPEP-----APTTPKEAPTTTPKEPAPT 282  
 | : | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 123 --DNKRAPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 180  
 QY 283 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KAPTTTPKEPAPTTPPTT 329  
 | : | | : | | | | | : | | | | | : | | | | | : | | | | | :

DB 181 APDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 240  
 QY 330 PKE--PAPTTKEP-----APTTPKEPAPTTPK-----KAPTTTPKEPAPTTPKEPAPT 376  
 | : | | : | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 241 APDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 300  
 QY 377 TKEPSTTPKKEPAPT--TKSAPTTTKEPAPT-----TKSAPTTTPKEPAPTTPK-----E 425  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 301 APDXRP--XPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 359  
 QY 426 PAPTTTPKEPAPTTPK-----KAPTTTPKEPAPTTPK-----EPAPTTPKKEPAPTTPKEPAPT 477  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 360 SAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVT 419  
 QY 478 TKEPAP-----TTPKKLTPTPEKLTAPTTPEKAPTTPEELAPTTPEEPPTTPPEPAPT 533  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 420 SAPDXRXPGSTAPXAHGVT-----APDXRXPGSTAPXAHGVTAPDXRXP-----PGST 471  
 QY 534 TPKA-----AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 588  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 472 APXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXP--XPGSTAPXAHGVTAPDXRXP-----APDXR 530  
 QY 589 TTP-----KKAPKEAPTTTKEPSTTSOKPAP--TTPKGTAPTTTPKEPAPTTP 636  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 531 TAPXAHGVTAPDXRXPGSTAP--XAHGVTAPDXRXPGSTAPXAHGVT-----APDXR 585  
 QY 637 KEAPAPTTPKG--TAPTTTPKEPAPTTP-----KKAPKEAPTTTGGPTSTSD 682  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 586 PXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAP--XAHGVTAPDX 644  
 QY 683 KPAP--TTPK-----ETAPTTTPKEPAPTTP-----KKAPTTPEPTTPPTSSTVSNP 726  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 645 RPXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXP-----TAPXAHGVTAP 702  
 QY 727 TTPKEPT-----HKSPDESTPELSAPPTKALENSPEKRGVPTTKTAA----- 772  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 703 DXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAPDXRXPGSTAPXAHGVTAP 762  
 QY 773 -TKPEMTTAKDK--TTERDLTTPETTTAARPKETATTEKTEESITNTTTQVTS 828  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 763 DXRXPGSTAPXAHGVTAPDXRXPST--APXAHGVTAPDXRXPGSTAPXAHGVT 820  
 QY 829 TTTQDTTPFKITTLTKTTLPKVTTTTKTTTTEIMKPEETAKPKDRATNSKATTPKQ 888  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 821 APDXRXPXG-----STAPXAHGVT-----SAPDXRXPSTAPXAHGVTAPDXRXP 868  
 QY 889 KPTKAPKKPTSTKRPKTPRVRKRPKTPTPPRKMTSTMEPLMPTSRIDAMIQTTTRNQT 948  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 869 -GSTAPXAHGVTAPDXRP--XPGSTAPXAHGVTSA--PDXRXP-----PGST 911  
 QY 949 -PNSKLVEVNRKSEDAAGEETHMLRPVHFPEVTPOMDYLPVRVNOGIIINPMLSD 1007  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 912 APXAHGVTAPDXRXPSTAPXAH-----GVTSAPDXRXPXGSTA-----PXAHG 957  
 QY 1008 ETNLCNGKRPVQGLTT-LRNGTLVAFRGHYFMMLSPFSPSPARRITTEYWGIPSPID 1062  
 | : | | | | | | | | | : | | | | | : | | | | | : | | | | | :  
 DB 958 VTSAPDXRXPXGSTAPXAHGVTSA-----PDXRXPXGSTAPXAHGVTAPDX 1003

RESULT 12  
 US-09-083-116-5  
 Sequence 5, Application US/09083116  
 Patent No. 6203795  
 GENERAL INFORMATION:  
 APPLICANT: CHAMON, Pierre  
 APPLICANT: KIENY, Marie-Paule  
 APPLICANT: LATHE, Richard  
 APPLICANT: HAREUVENT, Mara  
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
 TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS, L.L.P.

```

Qy 442 PA--PTTPKEPA--PTTPKKEAPPTTYYKRAPATAKKEAPPTTKEETAPTPKLPTTPREK 497
Db 309 PSTPTPYKAKPSYPPTYYKAKPSYPPTYKAKPSYP---PTTKAKPSYPPTYKAKPSYP--YK 362
Qy 498 LAPTPPE--KPAPTTPEELAPTTDEEPT--PTTPEEPA--PTTPKAAAPNPKKPAAPTPK 552
Db 363 AAPTPYSTYKAKPSYP---PTYKAKPSYPPTYKAKPSYPPTYK--AKPSYP---PTYKA 413
Qy 553 EPA-PTT-----PKKAPPTTKEETAAPT--PKGTAPTTLKCPADPTTPKKAPKELAPTTTK 605
Db 414 KPTPTSTYKAKPSYPPTYKPKISYPTPYKAKKPSYPTYYKAKSSYPTPYKAKPSYPPTYKA 473
Qy 606 EPTSTTSKPAAPTYYKGTAPPTPKKPAAPTTPKKEAPPTTPEKAPPTTGGTAAPTLKLEPA--PTTPKPK 663
Db 474 KPTPTSTYKAKPTYK--AKPSYP---PTYKAKPS-----YPTTKAKPSYPTPYKAP 521
Qy 664 AKKELAPTTTGGPTSTSDKPAAPPTTKEETAAPTTPKEPA--PTTPKKA-PTT-----PETP 716
Db 522 TYK-AKPTPYKAKPSYPPTYKAKPSYP---PTYKAKPSYPPTYKAKPTVPSYTYKAKPSYP 576
Qy 717 PPTTSVSTPTTPTKEPTTIHKSPDESTPELSAPPTPKALENSKEGVTYTPATPTPE 776
Db 577 PTYKAKPSYPT-----YKAKPSYPTTKAKPTTKA-----KPTYST--YKAKPS 620
Qy 777 MTTTAKOKTERDLRTTPTTTAAP 801
Db 621 YPTTHKAKPTYKAKPSYPTTHKAP 645

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6130081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBRÖEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHÉLE D.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; NUMBER OF SEQ ID NOS: 1998-12-23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; US-09-219-849-5

```

Qy	311	PTTPEKAPATT--KEPRTTKKBA--PTTKKAPATTKEBA-----	34
Db	187	P--PGAPGAGPGGSRDOPGAPCPAGPGSGRDGPGAGHAPGKAGHAPGAPGKAGH	244
Qy	350	-----PTAPK-KAPATTP--KEBAPTTKEBA-PTTKKEPSPTTKKAPATTTKS	395
Db	245	PAGPKGAGHAPGPKGAPGAPGPGSGRDGPGAPCPAGPAGPSSRDBG--PGAPG----	A 296
Qy	396	APTTTKEAPTTTKSAPTTPKESPSTTTKEBAPTTKKEBAPTT--KKAPATTKEBAP	453
Db	299	GGPGSRDBGP-----PGAPGAGPGGSRDBGP--PGAPGAPGPGSRDBGP--PGAPG	346
Qy	453	TP--KEBAPTTKKBA-----PTAPKKEBAPTTKETAPTTPKILPTTPEKLAPTT	502
Db	349	AGPSSRDBPGAPGAPGPGSGRDGPGAPGAPGAPGPGSGRDBP--PGAPGAPGPG	404
Qy	503	PEKAPAT-TPBELAPTTPEEPTPTTPEBPATTTKAA--PNTKEBAPTT--KEBAPT	557
Db	405	SRDPPGAPGAPGPGSGRDBP--PGAPGAGPGSGRDBPGAPCPAGPGSGRDBGP	463
Qy	558	TPKEBA-----PTTKET-APTTPKGT-----APTTKEBAPTTKKP-APKELA	600
Db	463	GAGGAPGKAGHAPGAPGAGHAPGAPGAGHAPGAPGAPGAPGSGRDGPGAPBPAG	522
Qy	601	PTTKKEPSTISDKAPATTPKGTA-PTTKKEBAPTT--KEBAPTTKGT-PTTLKEP	655
Db	533	PGSGRDBPGCA--PGAPPGSGRDBPGAPGAPGPGSGRDBPGAPGAPGPGSRDP	580
Qy	656	APTTPKKAP--KELAPTTTKGFTST--TSDKAPATTPKETAPTTKEBAPTTPKK	707
Db	581	GP--PGAPGAPGPGSGRDBPGAPGAPGAPGPGSGRDBPGAPGAPGPGSGRDBP--PGA	636
Qy	708	PAPTP--ETPPTTSEVSTPTTKKEPTTIHKSPEDESTPELSAETPKALBN-----S	758
Db	637	PGPAPGPGSRDBPGAPGAPGAPGPGSGRDBP--PGAPGAPGPGSRDBPGAPGAPGAP	695
Qy	759	PKBGPVTTKTPATK 774	
Db	696	SRDPPGAPGAGPAPK 711	

```

1  : APPLICANT VAN DEN BOSCH, TANJA J.
2  : TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
3  : TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
4  : FILE OF INVENTION: PREPARATION THEREOF
5  : FILE REFERENCE: 2728-2
6  : CURRENT APPLICATION NUMBER: US/09/219,849
7  : CURRENT FILING DATE: 1998-12-23
8  : NUMBER OF SEQ. ID NOS.: 50
9  : SOFTWARE: PatentIn Ver. 2.1
10 : SEQ ID NO 5
11 :
12 : LENGTH: 960
13 :
14 : TYPE: PRT
15 :
16 : ORGANISM: Artificial Sequence
17 :
18 : FEATURE:
19 :
20 : OTHER INFORMATION: Description of Artificial Sequence: Illustrative
21 :
22 : US-09-219-849-5

```

RESULT 11  
US-08-479-537A-5  
Sequence 5, Application US/08479537A  
Patent No. 5861381  
GENERAL INFORMATION:  
APPLICANT: CHAMBON, Pierre  
APPLICANT: KIENY, Marie-Paule  
APPLICANT: LATHE, Richard  
APPLICANT: HAREVENT, Marda  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE  
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,537A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 90/13101  
FILING DATE: 23-OCT-1990  
PRIOR APPLICATION DATA:

```

OY      696  TKEDP-----APTPKKPAPTPTPTPTTSEVSTPTTKKEPTTIHKSPDESTPELSAE 749
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      650  LPPIQKGNINIPSNIPSPDSEVEYPRPDNGENSNNTKSKKNI---PNEPIPSGDN 706
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      750  PTPKALENSPK-----EPGVPTTK 768
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      707  PYKGHERIKPRHRSNDYIYDNNVKNKNNKNDPEPIIPNNE 745
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT      8
PCT-US92-00018-2
Sequence 2, Application PC/TUS9200018
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenvit, Yupin
APPLICANT: Hedstrom, Richard
APPLICANT: Khumsmith, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene encoding
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: A. David Spevack
STREET: NMDC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00018
FILING DATE: 19920103
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avram D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-4033
TELEFAX: (301) 295-6759
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00018-2

```

Query Match	7.2%;	Score 488.5;	DB 5;	Length 826;
Best Local Similarity	26.4%;	Pred. No. 11e-24;		
Matches 153;	Conservative	54;	Mismatches 223;	Indels 149;
				Gaps 34;
Oy	233	TTTTIKSAPTTKEEPA-PTTTSAPTTKEEPA-PTTTKEP---	APTTPKEP-APTTPKE	285
	11::			
Db	273	TPCKVRDCQIPILPIVPIPKIPKPSNPDEEVPVNDPNDPNNNNNNNNNNNNNNNNNN		332
Oy	286	PAPTTKSAPTTTKEP-APTTPKKP-APTTPKEP-APTTPKEP-TPPTTKEEPAPTTKEEPA		341
	11::			
Db	333	PNNPNNDPNSPNNHKK		392
Oy	342	PTTTPKEEPAPTAPK-KPAPTTKEEPAPTTTPKEEPAPTTTKEPSTTTTKEEPAPTTTKSAPTT		399
	11::			
Db	393	RRNNKRRNNPKPKPKNNKPNKPNPNESNNKKPNP----	ESNNKKPNP----	NEPSN 443
Oy	400	TKEEPAPTTTSAPTTKEPSPITTTKEEPAPTTKEEPA-PTTPKKP-APTTPKEEPAPTTKE		457
	11::			
Db	444	PNNKPNP-----NEPSNPKNPNP-----EPLNPNEPSNPNEPSNPAPNSPNE--PSPNPE		492
Oy	458	PAPTTTKKAPAPTAPKEP-APTTPKEEPAPTTPKILITPTTKEKLAAPTTPKEKAPATTPPELAP		516

```

Db      493 PS-----NNEBSSNPNE--PSNPK-----PSNPNE--P 523
OY      517 TTPEEPTTPEEAPATTPKAAPTKEP-APTTPKEPATTTKEPATTTKATTP 575
Db      524 SNPNE--PLNNEP-----SNNEPSNPENPNPEB--PSNKE--PSNPNE-----564
OY      576 KGTAPTTLKEAPATTPKKAPKELAPTTTKPTSTSDKAPATTKGTAPTTKEAPATT 635
Db      565 -----PSNPEPNPEE--PSNPKRP-----SNPEPINPELNPKEPSNPEBN 606
OY      636 PREPATTPKGTATTTKEPATTPPKKAPKELAPTTTKGPTSTSDKAPATTKEPAT 695
Db      607 PKEPI-----NPEESNKEP-----INPEDENELIODEIEPRNDSNVPI 649
OY      656 TPKEP-----APTTPKKAPATTPETPEPTTSEVSTPTTKEPTTIHKSPESTPELSAE 749
Db      650 LPTPKGNNTIPSLPNPESSEVYTPRPNDNGENSTMTSKNNI---PNEPIPSPDN 706
OY      750 PTPKALENSPK-----EPGVPTTK 768
Db      707 PYKGEHERIPKPRHSNDYVYDNNVKNKKKDEELIPNNE 745

RESULT      9
5202236-13
: Patent No. 5202236
: Applicant: MUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
: SUSAN L.; MCCANDLIS, RUSS; WEI, TENA; FILPULA, DAVID
: TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
: PROTEIN
: NUMBER OF SEQUENCES: 39
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/528,762
: FILING DATE: 25-MAY-1990
: APPLICATION NUMBER: 82,456
: FILING DATE: 07-AUG-1987
: APPLICATION NUMBER: 933,945
: FILING DATE: 24-NOV-1986
: APPLICATION NUMBER: 650,128
: FILING DATE: 13-SEP-1984
: SEQ ID NO:13:
: LENGTH: 652
5202236-13

```

Query Match	7.0%;	Score 476.5;	DB 6;	Length 652;
Best Local Similarity	29.7%;	Pred. No.5,1e-24;		
Matches 221;	Conservative	71;	Mismatches 284;	Indels 169;
				Gaps 45;
QY 116	PKTTAKPIINPRSLPNDSQTSKETSIVNKEETVEKETTNNKOTSDGKEKTTSAKE	175		
DB 11	PKMTTPTTKPKKSYTP-----TYSKSPY-----KPKIN----	40		
QY 176	TOSIEKTSANDLAPTSKVIAKPPKAEKTTTGALTTPEKEPTTTKEA-----PASTTPE	231		
DB 41	-----YPTPKY-AKPS-----YPTPKPKKTPPTTKPKLTTPTTKPKP	79		
QY 232	PTPTTTKSAPT--TPKEAPPTTTSKAPTTKEAPPTTKKEAPPTTKEAPTTTKEA--P	288		
DB 80	SYPTTKSKPTTKPKLTTPTTKAKSSYPTTKPKKTTPTT--YKPLTTPTTKPKASYP	138		
QY 289	TTTTSAPTTPEKAPPTTKKAPAPT--TPKEAPPTT---PKESPPTTKKEAPPTTKEAPPT	343		
DB 139	PTTKPKKSYTP--PSYKTKKLTYPPTTKKLTYPPTTKPKPSYPSYKPNKTYPTTKPKLT	196		
QY 344	TPKEAPPTAKKPA--PTTKEKA--PTTKEAPPTTKKPSPTT-----KCAPPTTK	394		
DB 197	YP-----PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPLIAKSSYPTTK	252		
QY 395	SAPTTTKEA-----PTTTSAPTTPKESPPTTKKEA--PTTKEKA-----PTTKK	441		
DB 253	AKPTTKAKPKTYKAKPTYSKAKPTTP-----PTTKAKPSYPTTKAKPKTYKAKPKTYKAK	308		



```

QY 415 PKE-----PSPTT----- 422
DB 407 TSETESVIRKDEMCWLEKNGECEAKGATVGVIGKDRIGENGMATMIPNDTHFRFK 466
QY 423 -----TKEP-----APTPKE 433
DB 467 VADVGNITISRCRKAGKLEFPDRSLDFTIPVAGHNSCSIIIVGSGDKIHVSFGSKD 526
QY 434 PAPTTPKKEPAPTTPKE-----PAPTTPKKEPAPTTPKKEPAPTTPKKEP 475
DB 527 VSLIS---APIQSELEFNEVYCDCTAKYGAHSGYOTSADPVTYTTAKPPTTT--TGA 580
QY 476 PTTKETAPTTAPTTKLLPPTTPKELAPTTPEKPAPTTPEELAPTTPEEPTTPPEEPAPTTP 535
DB 581 PQQPTTYYTGSPPKPTTTTATTTT-----TLNPLITTTTQKPTTT-----TT- 627
QY 536 KAAANPTKREPAPTTPKREPAPTTPKREPAPTTPKATPTTKATPTTKREPAPTTPKKA 595
DB 628 --KVGCKPPIATTTTTLKPIYTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 680
QY 536 PK---ELAPTTKKEPTS-----TTSKREPAPTTPKG--TAPTTPKREPAPTTP 635
DB 681 PDICDIEITPIEKMLKQYRMIVDYNSGLLSDNDPEIGSQAGQIADTNSLEPVQTH 740
QY 636 PKEPAPTTPKGTAPTTTKE---PAPTTPKKAPELAPTTTKGPTSTTSKREPAPTTPKET 692
DB 741 KSTGLPIDMVGLEPDPKSGNLVHPYTNQTMGSLSVSYLAKNLVDDETYG--LPIDT 798
QY 693 APPTTPKREPAPTTPKREPAPTTPPTTPPTSEVSTPT-----TTKETPTTHKS----- 738
DB 799 LTGVPLDVSILIPN--PETSELEDPSIDELMNGTIAGIYSGISASELSOKALIDPA 856
QY 739 -----PDEST-----PELSAEPPTPKALENSPK 760
DB 857 TNNVYGEFGLINPATGVIMIPGLPSEHQFSPSEIEDGGIIPREVAANDKRLSLP- 915
QY 761 EPGVPTTKTPATKPEMTTAKDKTER-----DRTTPETTTAAPKM 803
DB 916 -PSVP-----ESLPE-----KDKKIDISIELMWDIESGRIGOVSKRPIPGSIAGDLP 963
QY 804 TKETATTEKTESKITAT--TOVSTTTODTTPPKITTTTTLAKVTTTKKTTTTE 862
DB 964 IMKPTOTDSVYKRPDPTGLPFPNPTGLHINLTNNNTDSSAAGVAKVAVSGITDN 1023
QY 863 IMNKP--EENAKPKD-----RATNSKATTPKQKPTKAPKPTSTKKKKTMP----- 907
DB 1024 VYGLPVEITGLPKDPSGDIPEPNTSTGELVDPSTGKPINNSTAGIYSGKGLPIEDENG 1083
QY 908 -----RVRKEXTTPTPKMTS--TMPE----- 927
DB 1084 NLFDPSNLPIDGNLQVNPETNSTVSGTTPKPKPIPVNGGVVPDEEAKDQADKG 1143
QY 928 -----LNPTRIAEAMLOTTTRPNOTPNSKLVENPKSEDAAGETPMLLRPHVFP 982
DB 1144 KQGLIVPTINSINKDPVTNIOYSTGTGNI---INP--ETGKVIYGLSLPESLAVPSFNP 1197
QY 983 EYTPD-----MDYLPVFNQGIINP-----MLSDENYICGK----- 1015
DB 1198 QQDDELIRKQVDVYTGPIYPSSTGEIIDPATKLPISGAVAGDEILLEVINTTTDEVYGLP 1257
QY 1016 -----PYDGLTTLRNGTLV 1029
DB 1258 IDELETGLPRDPVSGLPOLPNGTLV 1281

```

```

RESULT 7
US-07-638-431-2
Sequence 2, Application US/07638431
Patent No. 5198533
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charoenvalt, Yupin

```

```

APPLICANT: Hedstrom, Richard
APPLICANT: Khushmith, Srisin
APPLICANT: Rogers IV, William O.
TITLE OF INVENTION: Protective malaria sporozoite surface protein
TITLE OF INVENTION: Immunogen and gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: A. David Spevack
STREET: NMDC Building 1 T-12 National Naval
STREET: Medical Center
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20814-5044
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/638,431
FILING DATE: 19910110
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, Avrom D.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-4033
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-638-431-2

Query Match 7.2%; Score 488.5; DB 1; Length 826;
Best Local Similarity 26.4%; Pred. No. 1,1e-24;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

QY 233 TPTTISAPTTTKEPA-PPTTKSAPTTPKKA-PPTTKP---APTTPKEP-APTTPKE 285
DB 273 TCKVNDQDIPPIPVIPKIPKPSNDEPVNPNDPNPNPNPNPNPNPNPNPNPNPNPN 332
QY 286 PAPTTPKSAPTTPKEP-APTTPKPP-APTTPKEP-APTTPKEP-PTTTPKEPAPTTPKE 341
DB 333 PNNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 392
QY 342 PTTTPKEPAPTTPK--KAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 399
DB 393 RRPKRRNNPNKPKPNKPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 443
QY 400 TKEPAPTTPKSAPTTPKESPTTTPKKEPAPTTPKKEP-PTTPKKP-APTTPKEPAPTTPKE 457
DB 444 PKNPNP---NEPSNKNKNPN---EPLNPNESNPNPNPNPNPNPNPNPNPNPNPNPN 492
QY 458 PAPTTPKKEPAPTTPKKEP-APTTPKETAPTTPKLLPTTPTEKLAPTTPKEPAPTTPKE 516
DB 493 PS-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 523
QY 517 TTPPEEPTTPPEEPAPTTPKAAAPTTPKEP-APTTPKEPAPTTPKKEPAPTTPKKE 575
DB 524 SNPNP---PLNPNP---SNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 564
QY 576 KGTAPTTPKEPAPTTPKKAPELAPTTTKETSTSDKRAPTTTKGTAAPTTPKKEPAPTTP 635
DB 565 -----PSNPEPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 606
QY 636 PKEPAPTTPKGTAPTTPKKEPAPTTPKKAPELAPTTTKGPTSTSDKRAPTTTPKETAPT 695
DB 607 PKEPT-----NPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN 649

```

[illegible]

```

US-08-928-361B-6
: Sequence 6, Application US/08928361B
: Patent No. 6071518
:
: GENERAL INFORMATION:
:   APPLICANT: Petersen, Carolyn
:   TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
:   TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
:   TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
:   TITLE OF INVENTION: SPECIES INFECTIONS
:   NUMBER OF SEQUENCES: 30
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: PETERS, VERNY, JONES & BIKSA
:   STREET: 385 Sherman Avenue, Suite 6
:   CITY: Palo Alto
:   STATE: CA
:   COUNTRY: USA
:   ZIP: 94306-1640
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/928,361B
:   FILING DATE: 12-SEP-1997
:   CLASSIFICATION:
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 60/026,062
:   FILING DATE: 13-SEP-1996
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Verry, Hana
:   REGISTRATION NUMBER: 30,518
:   REFERENCE/DOCKET NUMBER: 480.76-1(HV)
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 650-324-1677
:   TELEFAX: 650-324-1678
:   INFORMATION FOR SEQ ID NO: 6:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 1721 amino acids
:     TYPE: amino acid
:     STRANDEDNESS:
:     TOPOLOGY: linear
:     MOLECULE TYPE: protein
:   US-08-928-361B-6

```

```

; Sequence 3, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; NUMBER OF SEQUENCES: 4
; TITLE OF INVENTION: CDNA and Related Products and Methods
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Plinnist & Michaels, P.C.
; STREET: 118 No. 6187558th Tloga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: RTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 786 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-3

Query Match          7.38; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.18; Pred. No. 2.2e-25;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;

QY 197 PTPKAEATTGPAATTTKEPTPTTPKPEASTTTPKKEPTTTIKSAPTTPKPKAPATTTKSAP 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 93 PGPRAE-TTQAPATNT-----QAPTTTQAPTTT-----QAPTTTQ--ATTTTTQAP 135

QY 257 TTPKEAPATTKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPK 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 TTTC--APTTCQ--ATTTCQ--APTTCQ--APTTCQ--APTTCQ--APTTCQ--APTTCQ--APTTCQ 183

QY 317 PAPTTPKEAPTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATT 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 -APTTCQAPTTCQAPTTCQAPTTCQAPTTCQAPTTCQAPTTCQAPTTCQAPTTCQAPTTCQAPTTCQ 241

QY 374 -----PTTKKEPSPTTPK- 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 ELLPNCSPADFIDILIPHDKYCNLFYQCSNGYTFQRCRPEGLYFPRVYQRCDSIPANVC 301

QY 387 -----EPAPTTT 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 DGEISAPRPVTEGNEDEDIDIGDLNCGRANFEIDMLLPHGNRCDDKYQCWGNLVER 361

QY 394 -----KSAPTTTKEAPATTTPKKSAPTTPKKEPSPTTKPE 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 CGAGTHSFELQOCNHIELVGLTLPGESESEVDEDA-CTGWCYCTPIETIEMERLPNCP 420

```

```

QY 427 APTT-----PKRP-----APTTPKKAPATT----- 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 ADFSIDHLLPHESDCGYLQCVHGOIARPCPGNLMHSPATOSCESEVTAAGCVFECDS 480

QY 447 ---PKRAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATT 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 NQCTSTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 533

QY 504 EKPAATPEELAPTTPEEPPTTPPEEPAPTTPKKAAPNT--PKRAPATTTPKKEAPATTTPKE 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 --PAPTAPTAAPTAAPTAAPESPTTVTP--PTAAPTAAAPTAVPEIPIVTV--SAPTAAPT 587

QY 562 PAPTTPKEAPTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTP 621
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 AAPTAAPTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 632

QY 622 GTAPTTPKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAP 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 TAAPTTP--AAPAPNTTVTPPTAAPTAAPTAAPTAAH-----APTAAAPTTPTS 677

QY 682 DKRAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPK 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 ---APTTPPEDDIDP--PLPNDPIP 699

RESULT 5
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LIECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GOT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5

Query Match          7.28; Score 489; DB 3; Length 1721;
Best Local Similarity 22.18; Pred. No. 2.4e-24;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

QY 70 KKNRTKKRPKPPVVDAGSLONGDEKVTPTSTTQHNKVSSTKITTAKNRPS 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 RSNETK---TTEPSANTYAGVYRSN-EKTTPEBANT---NELLVDPRT-----N 158

QY 130 LPPNSDTRKE-----TSLFVNKEETVETKETTTTNKOTSDGKEKTSK 174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 APCNSENSEFQGIIFDMGSKYIIPYTCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 208

QY 175 ETOSIEKTSARDLAPTSKVLAKPTPKAETTTKGALTTTPKEAPTTPTPKKEASTTPKEPTP 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240

QY 235 TTTSAPTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKSA 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 300

QY 295 PTPKKEAPATTTPKKEAPATTTPKKEAPTTPKKEAPTTPKKEAPATTTPKKEAPATTTPK 354

```

```
QY 477 -----TTPKETA-PTTP-----KKL 490
;      | | | | |
Db 842 AGQIADSNLFVQTHKSTGLPIDRWGLPDPKSGNLVHPTNQTMSGLSYLAANKL 901
;
QY 491 T-----PTPEKLAPTPE-----504
;      | | | | |
Db 902 TWDDETYGLPIDLGLYGLDPISLIPENPETGELFDPISDEIMNGTAGIVSGISASES 961
;
QY 505 ----KRAPTPPE-----LAPT-----PEEPTTPP--DEAPPTPKAA 539
;      | | | | |
Db 962 LLSQKAPIDPATNMVGEFGLNATGVMTPGSLGPSEQTPFSEIIDGGIIPPEVAA 1021
;
QY 540 PNTPKREPAPTPKEAPTPKE-----PA 563
;      | | | | |
Db 1022 ANADKKLSIPSVSPESIEPKDQKIDISBLMYDIESGRIGOVSKRPIPGSIAGDLNPI 1081
;
QY 564 PTPPKETAPTPKGAAPTLKEAPAPTPPKKPAKELAPTTKEPTSTSDKRAPTPPKGT 623
;      | | | | |
Db 1082 MKTPGTQDSVYGKPIDPTT---GLPFNP--PTGHLINPNNMTDSSFAGAYKVAVSNQI 1136
;
QY 624 APPT---KREPAPTPKREPAPTPPKGTAPTLKEAPATTP-----KKPAK-- 666
;      | | | | |
Db 1137 KTDNVYGLPVDLITGLPKDVSQDIPFNSTGELVDSGTGKPINNTAGIVSGKRLPPIE 1196
;
QY 667 ----ELAPTTTKGP-----TSTT-----SDKAPTPPKETAPTTKEPA-- 701
;      | | | | |
Db 1197 DENGNLDPSTKLPIDGNQLVNPENETNSVSGSTGSKTKPKGLPVNGGVVDEEAKDO 1256
;
QY 702 -----PTPKKRAPTPTPPETTPSEVSTPTTKE--PTTIHKS--PDESTPEL 746
;      | | | | |
Db 1257 ADKKGDLIVPINSINKDPVTNTQVSNNTGNIINPETGAVIRGSLGSLNYSFNPOQ 1316
;
QY 747 SAEPTRKALE---NSPKPEGVPTTKTPAATKPEMTTAKDKTERDLRTPETTAAPKA 803
;      | | | | |
Db 1317 TDELTKRPVDVITGLPDPSTGELIIDPATKLPISVAGSEIILEVINITTDEVTGLP-I 1375
;
QY 804 TKEPATTTKTESKITATTQVSTTQDTPPKITTKITTLAPVTTTKITTE- 862
;      | | | | |
Db 1376 DLEGLPBDP-----VSGLPOLPNGTLVD-----PSKKKRIPSHS 1411
;
QY 863 --INMKPEPAKPKDRATNSKATTPKPOK--TKAP--KKPTSTKPKTMPRVRKPKT-- 914
;      | | | | |
Db 1412 GFINTSGEQSHKDPSTG-----KPLDPNTGLPFDSDGSLINPETGDKLQSHSGTF 1465
;
QY 915 TPTPKMTSTMPELNPISRIAPAMIQTTTRPNQTPNSKLVEYNK--SEDA----- 963
;      | | | | |
Db 1466 MPVPGKPGENGIMTPQOILEAL-----NKLPTSEVNISPPSSDAVDPDRPTNTW 1518
;
QY 964 -----GGAEGETPHML-----RPHVMEVTPMDMDLPRVNOGIII 1001
;      | | | | |
Db 1519 NKISQOTQVDSKTIIPDSASVHITALGTPTQIDPTTGLP--SDPSTGLPFIQFENVLV 1576
;
QY 1002 NFMLSD-----ETNIC-----NGKPYDGLTTL---RNGLVV-AF 1031
;      | | | | |
Db 1577 DQGTGBOJIKGSVPYSLVYKERNIVTEAAYGLPVDKGTGFPIDPISLYLPAKNGELDPI 1636
;
QY 1032 RGHYF 1036
;      | |
Db 1637 SGKTYF 1641

RESULT 3
5202236-25
; Patent NO. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIODHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
```

```
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 25
; LENGTH: 744
5202236-25
```

```
Query Match 7.4%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 6.2e-26;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

116 PKITTAKINRPSLPMSDTSKESLTVNKEETVETKETTNNKQSTDEKERTSAKE 175
;      | | | | |
Db 24 PKMTYPPYKPKPSYP-----YKSKPY-----KPKIT----- 53
;
QY 176 TQSIKTSADLAPTSKYLAKTPKAEETTKGAPALTTPKEEPTTPKE---PASTTPKE 231
;      | | | | |
Db 54 -----YPTTYK--AKPS-----YPTTYKPKKTYPPYKPKLTPPYKPKP 92
;
QY 232 PPTPTIKSAPT--TKEPAPTTKSAPTTPKEAPATTPKEAPATTPKEAPATTPKEA--P 288
;      | | | | |
Db 93 SYPTTYKSKPYTKPKIYPPYTKAKSPYPTTYKPKKTYPPYTKKLTYPYTKAKSYP 151
;
QY 289 TTTKSAPTTPKEAPATTPKPAPT--TKEPAPTT---PKEPPTTPKEAPATTPKEP--- 340
;      | | | | |
Db 152 PTKPKPSYP--PSYTKTKITPTTYKPKLTPPYTKPKPSYPSYKPKTYPPYKPKLT 209
;
QY 341 -APTPKEPA--PTAKKPA--PTTPKEPA--PTTPKEPA--PTTPKEPS--PTTPKEPA 389
;      | | | | |
Db 210 YPTTYKAKSPYPTTYKAKSPYPTTYKAKSPYPTTYKAKSPYPTTYKAKSPYPTTYKAKS 269
;
QY 390 -PTTTSAPTTKEAPATTPKSAPTTPKEPSPTTKEPA--PTTPKEPA--PTTPKRP 442
;      | | | | |
Db 270 YPSTYKAKSYT---YPSYTKAKTYP-----PTYKAKSPYPTTYKAKPYTKAKPYTKAKP 322
;
QY 443 A--PTTPKEPA--PTTPKEPA--PTTPKPA--PTAKKPAPTTPKEPAPTTPKLTPTT 494
;      | | | | |
Db 323 SYPTTYKAKSPYPTTYKAKSPYPTTYKAKSPYPTTYKAKSPYPTTYKAKPYTKAKPYPS 382
;
QY 495 PEKLAPTPE--KRAPTPBELAPTPPEPT--PTPEEPA--PTPKA--AAPNT-----P 543
;      | | | | |
Db 383 TYKAKSPYPTTYKAKBSYP---PTYKAKSYPTTYKAKSPYPTTYKAKPYTKAKPYSTYKAKP 438
;
QY 544 KEAPATTPKEAPATTP---KREPAPTTKETAPTT--PKGTAPTLKEAPATTPKPAK 597
;      | | | | |
Db 439 SYPTTYKPKISYPTTYKAKSPYTYKAKSSYPTTYKAKSPYTYKAKSSYPTTYKAKP 498
;
QY 598 ELAPTTKEPTSTSDKRAPTTKGTAPTPKREPAPTTPEAPATTPKGAAPTLKEPA- 656
;      | | | | |
Db 499 SYPTTYKAKPLYSYTYKAKPTYK--ARPSY---PTYKAKPS-----YPTTYKAKPSV 546
;
QY 657 -PTPKKPAPELAPTTTKGPT--STSDKPAVTPKETAAPTPPKREPAPTTPKPA--PT 711
;      | | | | |
Db 547 PPTTKAPPTYK--AKPTYKAKPTYSTYKAKPSYPTTYKAKPSYP---PTYKAKPSYPT 601
;
QY 712 ---TPEPTPTTSEVSTPTTKE---PTTIHNSPD-----ESTPELSAAPT--PKALENS 758
;      | | | | |
Db 602 YKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPYTKAKPYKAPTYPSYTKAP 661
;
QY 759 PKEGVPTTKTPAA-----TKPEMTTAKDKTERDLRTTPETTTAAP-----KMT 804
;      | | | | |
Db 662 PSYP--PTYKAKPSYPTTYKAKPSYPTTYKAKPYTKAKPYTKAKPSYPTTYKAKPYTKAKP 719
;
QY 805 KETATTTKTESKITATT 824
;      | | | | |
Db 720 YKAKPTYSTYKAKPTYST 739
```

```
RESULT 4
US-09-103-429A-3
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:33 ; Search time 61.21 Seconds  
(without alignments)  
466,904 Million cell updates/sec

Title: AA3  
Perfect score: 6814  
Sequence: 1 MAWKTPYLLLLSVFYIQ.....ARAITRSGTSLSKWTNCP 1270

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

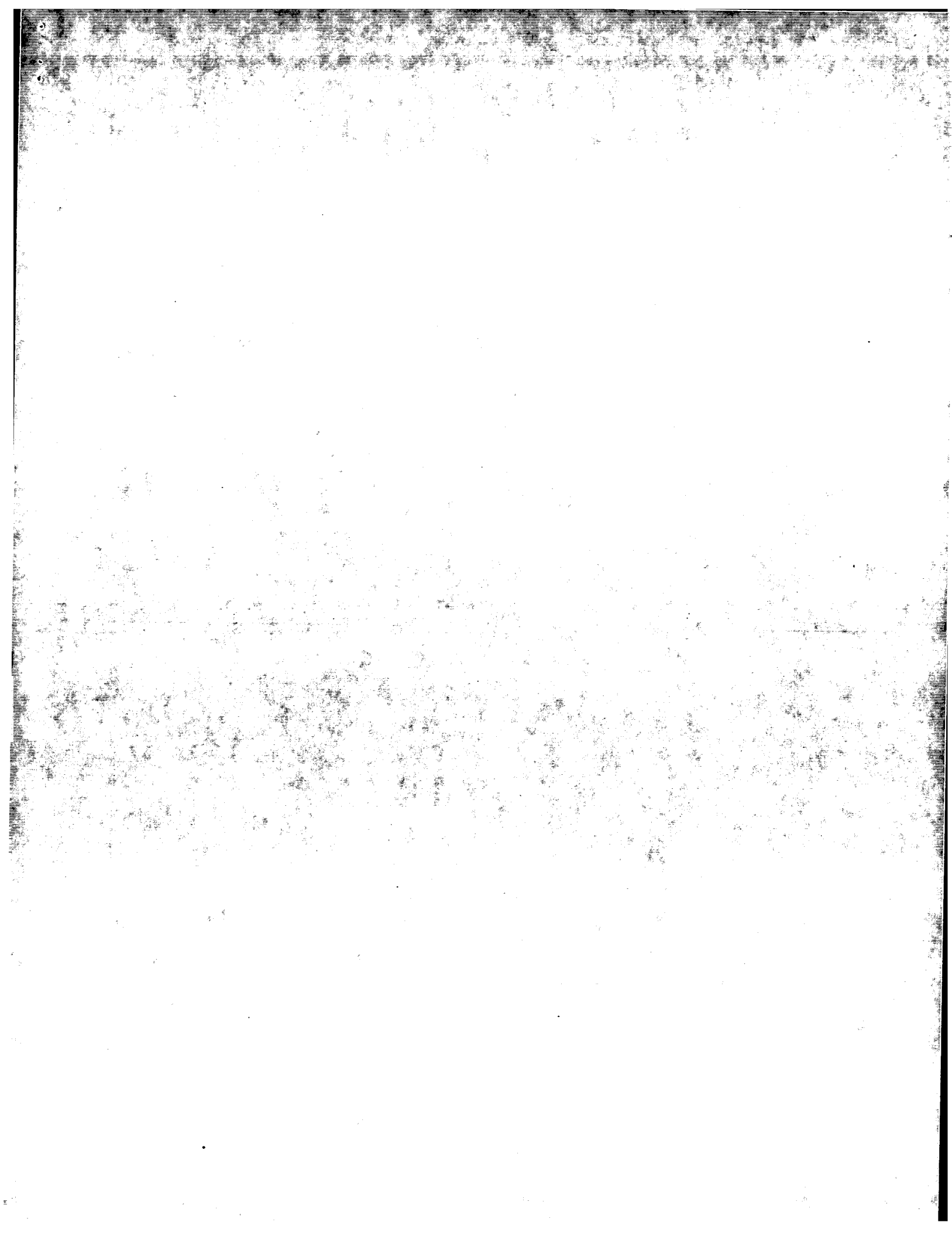
Result No.	Score	Query Match	Length	ID	Description
1	522	7.7	805	US-09-103-429A-4	Sequence 4, Appli
2	508	7.5	1837	US-08-928-361B-5	Sequence 5, Appli
3	506.5	7.4	744	5202236-25	Patent No. 5202236
4	498.5	7.3	786	US-09-103-429A-3	Sequence 5, Appli
5	489	7.2	1721	US-08-700-651-5	Sequence 6, Appli
6	489	7.2	1721	US-08-928-361B-6	Sequence 6, Appli
7	488.5	7.2	826	US-07-638-431-2	Sequence 2, Appli
8	488.5	7.2	826	PCR-US92-00018-2	Sequence 2, Appli
9	476.5	7.0	652	5202236-13	Patent No. 5202236
10	452	6.6	960	US-09-219-849-5	Sequence 5, Appli
11	424.5	6.2	1867	US-08-479-537A-5	Sequence 5, Appli
12	424.5	6.2	1867	US-09-083-116-5	Sequence 5, Appli
13	424.5	6.2	2035	US-08-479-537A-2	Sequence 2, Appli
14	424.5	6.2	2035	US-09-083-116-2	Sequence 2, Appli
15	419.5	6.2	2476	US-08-276-967-2	Sequence 2, Appli
16	419.5	6.1	1185	US-09-041-886-23	Sequence 23, Appli
17	417	6.1	829	US-08-642-255-132	Sequence 132, Appl
18	417	6.1	829	US-08-397-633A-33	Sequence 53, Appli
19	417	6.1	829	US-08-175-155-68	Sequence 68, Appli
20	417	6.1	837	US-08-477-509B-103	Sequence 103, App
21	417	6.1	837	US-08-642-255-101	Sequence 101, App
22	417	6.1	837	US-08-707-237A-75	Sequence 75, Appl
23	417	6.1	837	US-08-482-085B-103	Sequence 103, Appl
24	417	6.1	837	US-08-397-633A-50	Sequence 50, Appli
25	413.5	6.1	907	US-08-783-774-2	Sequence 2, Appli
26	413.5	6.1	907	PCR-US95-04611A-19	Sequence 19, Appli
27	404.5	5.9	408	US-07-609-716-65	Sequence 65, Appli

28	404.5	5.9	408	US-08-475-411A-65	Sequence 65, Appli
29	404.5	5.9	408	US-08-478-029A-65	Sequence 65, Appli
30	398.5	5.8	682	US-08-642-255-126	Sequence 126, App
31	398.5	5.8	682	US-08-397-633A-36	Sequence 36, Appli
32	381	5.6	1848	US-08-296-791-6	Sequence 6, Appli
33	381	5.6	1848	PCR-US95-10661A-6	Sequence 6, Appli
34	376	5.5	960	US-09-219-849-6	Sequence 6, Appli
35	375.5	5.5	1231	US-08-904-263A-4	Sequence 84, Appli
36	374.5	5.5	1537	US-08-325-267A-2	Sequence 2, Appli
37	371	5.4	761	US-08-707-237A-84	Sequence 114, App
38	371	5.4	762	US-08-642-255-114	Sequence 26, Appli
39	371	5.4	762	US-08-397-633A-26	Sequence 62, Appli
40	370.5	5.4	1064	US-08-642-255-62	Sequence 120, App
41	368	5.4	762	US-08-642-255-120	Sequence 31, Appli
42	368	5.4	762	US-08-397-633A-31	Sequence 28, Appli
43	367.5	5.4	1187	US-08-320-559-28	Sequence 28, Appli
44	367.5	5.4	1187	US-08-545-860D-28	Sequence 28, Appli
45	367.5	5.4	1187	PCR-US94-04496-28	Sequence 28, Appli

## ALIGNMENTS

```
RESULT 1
Sequence 4, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: cDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Brown, Piniel & Michaels, P.C.
STREET: 118 No. 6187558th Tloloa
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 805 amino acids
TYPE: amino acid
STRANDNESS: single
MOLECULE TYPE: protein
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-4
```

Query Match 7.7%; Score 522; DB 4; Length 805;  
Best Local Similarity 27.8%; Pred. No. 6.5e-27;







QY 508 PTTPEELAPTTPEPTPTTPEEPATTTKAAAPNTPKPEAPTTKPEAPTTKPEP--APT 565  
 DB 1778 PTPSP-NVSPSPSPSY-SPTSPSY-SPTSP-SYSPSPSPR-----YTPQSTP-YTPSPSPSYSPS 1828  
 QY 566 TPKETAPTTKPGTAPTTLKKEPAPTTPK-KPAKELAPTTTKEPTSTSDKPAPTTPKGA 624  
 DB 1829 SP-SYSPSPSPSYT-----PTSPSPSPSPSPSYTPASPK--YSPSPSPSPSPSPK--YS 1875  
 QY 625 PTTPEKAPTTKPEAPTTKPGTAPTTLKKEPAPTTKPKAPKELAPTTTGTPT-STTSDK 683  
 DB 1876 PTPSTY-SPTTPK-KPTSP--YTPSPSPSYTPSPK-----YSPSPSP--PTSPSPSPK 1922  
 QY 684 PAPTTPKEMAPTTKPEP--APTTPKPAPTTPTTPTTSEVSTPTTKEPTTHKSPDE 741  
 DB 1923 YSPSPSP--YTPSPSPSPSPSYTPSPSPK--YTPSPSPSPSPK-----YTPSPSPSP 1965  
 QY 742 STPE 745  
 DB 1966 SDEP 1969

## RESULT 15

NEFH\_RAT STANDARD: PRT: 831 AA.

AC P16884: 063368: 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)  
 GN NEFH OR NEFH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI-TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=89065087; PubMed=3143606;  
 RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;  
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).  
 RL Identification of putative phosphorylation sites.";  
 RL FEBS Lett. 241:213-218(1988).  
 RN [2]  
 RP SEQUENCE OF 37-831 FROM N.A.  
 RX MEDLINE=88309090; PubMed=2457365;  
 RA Dautigny A., Phan-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,  
 RA Jolles P.;  
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and  
 RL in situ detection.";  
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).  
 RN [3]  
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.  
 RX MEDLINE=87080760; PubMed=2878828;  
 RA Robinson P.A., Wion D., Anderson B.H.;  
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide  
 RL (NF-H).";  
 RL FEBS Lett. 209:203-205(1986).  
 RN [4]  
 RP SEQUENCE OF 318-831 FROM N.A.  
 RX MEDLINE=89184647; PubMed=2928342;  
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldhaber D.,  
 RA Smolowitz M., Carroll Z., Emanuel B.S., Breiner J., Rubin L.;  
 RT "Cloning of a cDNA encoding the rat high molecular weight  
 RL neurofilament peptide (NF-H): developmental and tissue expression in  
 RT the rat, and mapping of its human homologue to chromosomes 1 and  
 RL 22.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).  
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT  
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.

CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETT K-S-P, NEFH IS  
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS  
 CC THOUGHT THAT PHOSPHORYLATION OF NEFH RESULTS IN THE FORMATION OF  
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE  
 CC OF AXONAL CALIBER.  
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING  
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE  
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND  
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783  
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M37227; AAA41693.1; ALT\_FRAME.  
 CC EMBL: X13804; CAA32038.1; ALT\_FRAME.  
 CC EMBL: M21964; AAA41695.1; -.  
 CC EMBL: J04517; AAA41692.1; -.  
 CC PIR: A30796; A30796.  
 CC PIR: A25649; A25649.  
 CC PIR: B25649; B25649.  
 CC PIR: S02003; S02003.  
 CC InterPro: IPR001664; IF.  
 CC Pfam: PF00038; filament. 1.  
 CC DR PROSITE: PS00226; IF; 1.  
 CC DR Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.  
 CC KW NON\_TIR 1  
 CC FT DOMAIN 1  
 CC FT 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.  
 CC FT CONFLICT 164 164 L -> I (IN REF. 2).  
 CC FT CONFLICT 185 185 I -> S (IN REF. 2).  
 CC FT CONFLICT 193 193 L -> T (IN REF. 2).  
 CC FT CONFLICT 199 199 M -> T (IN REF. 2).  
 CC FT CONFLICT 346 346 K -> T (IN REF. 1).  
 CC FT CONFLICT 373 373 A -> V (IN REF. 1 AND 4).  
 CC FT CONFLICT 482 482 G -> E (IN REF. 2 AND 4).  
 CC FT CONFLICT 485 485 P -> S (IN REF. 2 AND 4).  
 CC FT CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).  
 CC FT CONFLICT 591 591 P -> T (IN REF. 2 AND 4).  
 CC FT CONFLICT 727 727 A -> V (IN REF. 4).  
 CC FT CONFLICT 757 759 AAP -> GST (IN REF. 4).  
 CC FT CONFLICT 769 769 T -> L (IN REF. 2).  
 CC FT CONFLICT 775 775 R -> P (IN REF. 2 AND 4).  
 CC SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

Query Match 7.3%; Score 497; DB 1; Length 831;  
 Best Local Similarity 27.6%; Pred. No. 6.1e-14;

Matches 195; Conservative 59; Mismatches 307; Indels 146; Gaps 35;

QY 102 PDRTSTQHNKVNSTPTITAKINPRPSLSPNSDSKESLIVNKEETVETKETTNTKQ 161  
 DB 197 PSMST--HIKVSSEKIKV-----KSEKTVIEEDQVEIOVEEYVEETED 242  
 QY 162 TSTDGKETTSAKETQSIETSAKDLAPTSAKYLAFTPAETTTGPAULTTPKPEP---T 217  
 DB 243 KKAOGGEEDAEDEGEAEATTS-----PPAEAAAP-----EKETSPKKEEKAESA 293  
 QY 218 PTTPKPEA-STTPKPEPTTTTISAPTTTKEPA-----PTTTSAPTTPKEPA---PTTTK 268  
 DB 294 PAEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 352  
 QY 269 EP-----APTTPKEPA-----PTTTTKEPA-----PTTTTSAPTTPKPEA---PTTTK 310  
 DB 353 SPGEAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPAFAKSPA 411  
 QY 311 --PTTPKPEA-----PTTTPKEP-----TPTTPKEPAPTTKEP---APTTPKEPA---PTA 352

SEQUENCE 1970 AA: 217205 MW: 6876FC25692A657E CRC64;

Query Match 7.4%; Score 503.5; DB 1; Length 1970;  
Best Local Similarity 33.1%; Pred. No. 6.8e-14;  
Matches 181; Conservative 101; Mismatches 167; Indels 95; Gaps 47;

QY 216 PPTTTPKEPASTTPKE-PTPTTTSAPTTPKEPAPTTTTSKAPTTPKE-----PAPTTPKE 269  
DB 1507 PSPMGISIPAMTWMGATATGAGMSVSGMTPGAAGSPSAAASDAGSPGSPAWMS 1566  
QY 270 PAPTTPKEPAPTTPKEPAPTTTTSAPTTPKEPAPTTTPKKAAPTTPKEP--APTTPKEPTP 327  
DB 1567 PTPSPSGSPSSPYIPSPGAGMSPSYSPTSPA-YEPSPGCGYTQSPSYSPSPSY-SP 1624  
QY 328 TTPKEPAPTTPKEPAPTTPKKAAPTTPKEPAPTTTPKEPAPTTTPKEPPTTPKE 387  
DB 1625 TSPSY-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-----SPTSPSY 1671  
QY 388 PAPTTPKSAPTTPKEPAPTTTTSAPTTPKEPPTTPKEPAPTTTPKEPAPTTP 447  
DB 1672 -SPTSPSYSPSPSY-SPTSPSYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1725  
QY 448 KEPAAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKA 507  
DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPNY-SPTSPNYT-PTSP-SYSTSP-SYSTSPNY-T 1777  
QY 508 PTPPELAPTPPEPTTPPEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTPKEP--APT 565  
DB 1778 PTSP-NYSPISPSY-SPTSPSY-SPTSP-SYSTSP-----YTQSPPT-YTPSPSYSPS 1828  
QY 566 TPKEAPTTPKGTAPTTPKEPAPTTPK-KAPKELAPTTPKEPTSTSDKAPTTPKGT 624  
DB 1829 SP-SYSTSPSKYT-----PTSPSYSPSPSEPTTPSPK--YSTSPKSPSPK-YT 1875  
QY 625 PTPPEPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKA 683  
DB 1876 PTSPNY-SPTTPKY-SPTSP--TSPSTSYVYTPSPK-----YSTSP--PTSPSTSPK 1922  
QY 684 PAPTTPKETAAPTTPKEP--APTTPKKAAPTTPPEPTTPSEVSTTPKPTTPKSPDE 741  
DB 1923 YSTSP-TYSPSTSPKSTYSPSTSPSY-SPTSP-----TYSLSPAL-----SPD 1965  
QY 742 STPE 745  
DB 1966 SDEE 1969

RESULT 14  
RPTL\_MOUSE STANDARD: PRT: 1970 AA.  
ID RPTL\_MOUSE  
AC P08775;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPL1).  
GN POLKA OR RPO2-1 OR RPII25.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SOURCE FROM N.A.  
RX MEDLINE=87280135; PubMed=3038894;  
RA Ahearn J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;  
RT Cloning and sequence analysis of the mouse genomic locus encoding  
RT the largest subunit of RNA polymerase II.;  
RL J. Biol. Chem. 262:10695-10705(1987).  
RN [2]  
RP SEQUENCE OF 1587-1970 FROM N.A.  
RX MEDLINE=66068017; PubMed=2999785;  
RA Corden J.L., Cadena D.L., Ahearn J.M. Jr., Dahmus M.E.;  
RT A unique structure at the carboxyl terminus of the largest subunit

RT of eukaryotic RNA polymerase II.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).  
RN [3]  
RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.  
RX MEDLINE=92178992; PubMed=1542581;  
RA Winzerleth M., Acker J., Vicaire S., Vigneron M., Kedinger C.;  
RT "Complete sequence of the human RNA polymerase II largest subunit.";  
RL Nucleic Acids Res. 20:910-910(1992).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
CC RNA(N).  
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- PTM: THE TANDEM 7 RESIDUE REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
CC THE PHOSPHORYLATION ACTIVATES POL2.  
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
CC III FOR 5S AND TRNA GENES.  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
CC -----  
DR EMBL: M12130; AAA40071.1;  
DR EMBL: M14101; AAA40071.1; JOINED.  
DR PIR: A28490; A28490.  
DR MGD: MGI:98086; Rpo2-1.  
DR InterPro: IPR006984; Rpo2-1.  
DR InterPro: IPR007022; RNA\_POL\_A.  
DR InterPro: IPR002879; RNA\_POL\_A2.  
DR Pfam: PF00623; RNA\_POL\_A; 1.  
DR Pfam: PF01854; RNA\_POL\_A2; 1.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 42.  
DR TRANSFERASE: DNA-directed RNA polymerase; transcription; zinc; Repeat;  
KW DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.  
FT ZN-FING 71  
FT DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
FT CONFLICT 1498 1498 P->R (IN REF. 1 AND 2).  
FT CONFLICT 1499 1536 MISSING (IN REF. 1 AND 2).  
FT SEQUENCE 1970 AA: 217175 MW: 7076F38F092A657E CRC64;

Query Match 7.3%; Score 498.5; DB 1; Length 1970;  
Best Local Similarity 33.1%; Pred. No. 1.1e-13;  
Matches 180; Conservative 101; Mismatches 168; Indels 95; Gaps 47;

QY 216 PPTTTPKEPASTTPKE-PTPTTTSAPTTPKEPAPTTTTSKAPTTPKE-----PAPTTPKE 269  
DB 1507 PSPMGISIPAMTWMGATATGAGMSVSGMTPGAAGSPSAAASDAGSPGSPAWMS 1566  
QY 270 PAPTTPKEPAPTTPKEPAPTTTTSAPTTPKEPAPTTTPKKAAPTTPKEP--APTTPKEPTP 327  
DB 1567 PTPSPSGSPSSPYIPSPGAGMSPSYSPTSPA-YEPSPGCGYTQSPSYSPSPSY-SP 1624  
QY 328 TTPKEPAPTTPKEPAPTTPKKAAPTTPKEPAPTTTPKEPAPTTTPKEPPTTPKE 387  
DB 1625 TSPSY-SPTSPNYSPTSPSY-SPTSPSY-SPTSPSY-SPTSPSY-----SPTSPSY 1671  
QY 388 PAPTTPKSAPTTPKEPAPTTTTSAPTTPKEPPTTPKEPAPTTTPKEPAPTTP 447  
DB 1672 -SPTSPSYSPSPSY-SPTSPSYSPSPSY-SPTSPSY-SPTSPSY-SPTSPSY-SPTSP 1725  
QY 448 KEPAAPTTPKEPAPTTPKKAAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKA 507  
DB 1726 SY-SPTSPSY-SPTSPSY-SPTSPNY-SPTSPNYT-PTSP-SYSTSP-SYSTSPNY-T 1777

FT DISULFID 374 391 BY SIMILARITY.  
 FT DISULFID 526 552 BY SIMILARITY.  
 FT DISULFID 536 551 BY SIMILARITY.  
 FT DISULFID 546 563 BY SIMILARITY.  
 FT DISULFID 573 599 BY SIMILARITY.  
 FT DISULFID 583 598 BY SIMILARITY.  
 FT DISULFID 593 610 BY SIMILARITY.  
 FT DISULFID 621 647 BY SIMILARITY.  
 FT DISULFID 631 646 BY SIMILARITY.  
 FT DISULFID 641 658 BY SIMILARITY.  
 FT VARIANT 276 276 K -> E.  
 FT VARIANT 354 354 K -> R.  
 FT VARIANT 415 415 T -> A.  
 SO SEQUENCE 662 AA; 67774 MW; F085277F1ED2FD40 CRC64;

Query Match 7.5%; Score 509.5; DB 1; Length 662;  
 Best Local Similarity 28.1%; Pred. No. 1.6e-14;  
 Matches 223; Conservative 48; Mismatches 276; Indels 247; Gaps 30;

QY 157 TTKKOTSDCKEKTAKETOSTIEKTSANDLAPTSKVLAKPTPKAETTTKGPALTTPKEP 216  
 DB 3 TTAATAAAGCKDTTAAEGSAAAEKTA-----AGEVSAPTT-AAVATGEDATT----- 51  
 QY 217 TPTPKPEASTPKEPTTTSAPTTKPEAPTTKSAPTTKEAPTTTKEAPTTTKEAPTTK 276  
 DB 52 -----AATAAATTAAAGAPTTTAPTTAAGKAPTTAATAATAPTTAAGAPTTAT 103  
 QY 277 EPAPTTKKEAPTTTKS-APTTKPEAPTTTPKPAPTTKEAPTTTKEPTTTPKEPA- 334  
 DB 104 GKAPATAAAPVPTTAAKAPTTAATAHSTAAAPTTAASAKKESSTSSSEEHCH 163  
 QY 335 -PTKEAPTT--TPKEAPTTAPKAPTTK-----EPAPTTKPEAPTTTKE 379  
 DB 164 VPKSKREKSGSGITKQ---CKKKNCCEFDKGGHGHCHFKPKGSHHEHTTTTK- 218  
 QY 380 PPTTPKPEAPTTTKSAPTTKPEAPTTKSAPTTKEPSTTKEAPTTTKEAPTTTKEAPTT 439  
 DB 219 -----APTTQIATTTT--TPPTT-----TTTKAPTT----- 245  
 QY 440 KKPAPTTKPEAPTTKPEAPTTTKKAPTTAKPEAPTTPEAPTTKPLTTPTEKLA 499  
 DB 246 -----TTTKATPTT-----TTTKATTTT-----TPPT----- 270  
 QY 500 PTPPKAPTTPEELAPTTPEEPPTTPEEPAPTTPKAAANTKEAPTTKPEAPTTTP 559  
 DB 271 -----TTTKATTTPTTTTTT----- 289  
 QY 560 KEPATTTKEAPTTPKATPTTLEKAPTTPKKPAKELAPTTTKEPTSTSDKPAPTT 619  
 DB 290 -----TTTKATTTTTSGECKMEPSK-----REDCGSGITTESOCR 328  
 QY 620 PKG-----TAPPT-----PKPEAPTTKPEAPTTPKGNAPTTLEKAPTTKPAPELA 669  
 DB 329 TKGCDFSDSIQTKWCFYTSQVADCKVEFSQRVDCRGIT---ADQCRKKKCCPDSS 384  
 QY 670 PTTTKGPTSTSDKPAPTTKEAPTTTPKPEAPTTPKKAPPTPEPTTPPTSEVSTPTT 729  
 DB 385 ISGKWCFSYSOVA--TKTTTPTTPTTPTTPTTPTTPTTPTT-----TTTTTTTTT 436  
 QY 730 KEPTTIHKSPESTPELSAEPTEKALNSPKREVPTTKPAATKPEMTTAAKDTERD 789  
 DB 437 TTTT-----TKA-----TTTTPTTPTTPTTPTT-----KAT----- 463  
 QY 790 LRTPEPTTAPKMTKEATTEKTESKTTATTTQVSTTTTODTTPPKITTLTKTTTLP 849  
 DB 464 -TTTTPTTTP-----TTTTKAT-----TTTTTTTTTTTTTTTTTTTTTTTTT 510  
 QY 850 KVTTKTTTITTEIMNKPEETAPKOKATNSKATTPKQKPTAAPKPKSTKPK-----T 905  
 DB 511 TTTTATTTTTTSGCKME---PSKADCGYGTESOCRSKGCDFSDSIQTKWCFYS 566  
 QY 906 MPVRKRPKPTTPR 919

DB 567 LPQVADCKVAPSSR 580  
 RESULT 13  
 ID REP1\_HUMAN STANDARD; PRT; 1970 AA.  
 AC P24928;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).  
 GN POLR2A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92178992; PubMed=1542581;  
 RA Wintzerith M., Acker J., Vicaire S., Vigneron M., Kedinger C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit.";  
 RL Nucleic Acids Res. 20:910-910(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95347616; PubMed=7622068;  
 RA Mita K., Tsuji H., Moriyama M., Takahashi E., Nenoi M.,  
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;  
 RT "The human gene encoding the largest subunit of RNA polymerase II.";  
 RL Gene 159:285-286(1995).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
 CC  
 DR EMBL: X63564; CAA45125.1; -  
 DR EMBL: X74874; CAA52862.1; -  
 DR EMBL: X74873; CAA52862.1; JOINED.  
 DR EMBL: X74872; CAA52862.1; JOINED.  
 DR EMBL: X74871; CAA52862.1; JOINED.  
 DR EMBL: X74870; CAA52862.1; JOINED.  
 DR PIR: S21054; S21054.  
 DR MIM: 180650; -  
 DR InterPro: IPR000684; RNA\_polII\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 43.  
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 71 87  
 FT DOMAIN 1590 1958 C2H2-type (POTENTIAL).  
 FT CONFLICT 1067 1067 W -> L (IN REF. 2).  
 FT CONFLICT 1449 1449 D -> Y (IN REF. 2).







TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>  
 CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
 CC -----  
 DR EMBL: M24496; AAA39813.1;  
 DR EMBL: M23349; AAA39813.1; JOINED.  
 DR EMBL: M24494; AAA39813.1; JOINED.  
 DR EMBL: M24495; AAA39813.1; JOINED.  
 DR EMBL: M35131; AAA39809.1; ALT\_FRAME.  
 DR PIR: J0368; QFMH.  
 DR PIR: A4378; A43778.  
 DR MGI: 97309; NFH.  
 DR InterPro: IP001664; IF.  
 DR Pfam: PF00038; Filament; 1.  
 DR PROSITE: PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;  
 KW Repeat.  
 FT DOMAIN 1 97 HEAD.  
 FT 98 408 ROD.  
 FT 409 1087 TAIL.  
 FT 436 517 GLU-RICH (ACIDIC).  
 FT 519 886 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.  
 FT 887 1087 GLU/LYS-RICH.  
 FT 98 129 COIL 1A.  
 FT 130 141 LINKER 1.  
 FT 142 239 COIL 1B.  
 FT 240 261 LINKER 12.  
 FT 262 283 COIL 2A.  
 FT 284 287 LINKER 2.  
 FT 288 408 COIL 2B.  
 FT 133 133 K -> QA (IN REF. 2 AND 3).  
 FT 133 133 A -> AR (IN REF. 2 AND 3).  
 FT 199 199 S -> T (IN REF. 2 AND 3).  
 FT 281 281 L -> G (IN REF. 2 AND 3).  
 FT 492 492 P -> PREAKSP (IN REF. 3).  
 FT 551 551 MISSING (IN REF. 3).  
 FT 689 712 G -> A (IN REF. 3).  
 FT 714 714 V -> M (IN REF. 2 AND 3).  
 FT 814 814 T -> N (IN REF. 2 AND 3).  
 FT 843 843  
 SQ SEQUENCE 1087 AA; 116612 MW; 57BAC76A38EDICB9 CRC64;

Query Match 7.8%; Score 533; DB 1; Length 1087;  
 Best Local Similarity 27.5%; Pred. No. 2,7e-15;  
 Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;

QY 86 DEASGIDNGFKY-----TPDSTTQHNKVNSTSPRTTAKPILNPRSLPNSDTSKET 140  
 DB 404 EECIGIGSPFSLTEGLPIPIST--HIKVKSEEMIKVE-----KSEKET 449  
 QY 141 SLTVNKETVETKETTNTKOTSDGKEKTSKETSISIKTSKADLAFTSKVLAPTRK 200  
 DB 450 VIVGGEERIRVEGVTEEDKEAGOGEEAEERGEKEEELAAATSPAAEAASPEKE 509  
 QY 201 AETTKAPALTTRKEPTPTTKPEBASTTPKEPTTITKSAPTTKPEBAPTTSAPTTPK 260  
 DB 510 TKSIVKKEEAKSPGKSPGAKSPA---EAKSPGKKS--PGGAKSPGKSPGAKSPAEKSPA 564  
 QY 261 EP-APTTKAPATTPEAPATTKEP---APTTKSAPTTKPEA---PTTPKPPA- 310  
 DB 565 EKPSPAAPAKSPA--EKSPSA--TVKSPGKSPGKSPGKSPGAKSPA---PTTPKPPA- 619  
 QY 311 ---PTTPKPPA---PTTPKPPA---PTTPKPPA---PTTPKPPA---PTTPKPPA- 355  
 DB 620 AKSPAAPAKSPAAPAKSPAATYKSPGKSPGKSPGAKSPA---EAKSPAAPAKSPAAPAEVKS 676

QY 356 P-----APTTKPEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 407  
 DB 677 PGEAKSPAEPKSPAPAA--KSPA-----EVKSPAEPKSPGKSPGAKSPA--PAVNSPAEAK 729  
 QY 408 TKSAPTTPKE-PSPTTKEPA-----PTTPKPEBAPTTPKPAPTTPKPEBAPTTPKEBAPTTP 462  
 DB 730 SPAAKSPGKSPGKSPGKSPGAKSPAEPKSPGKSPGAKSPAEPKSPGKSPGAKSPAEPKSPG 789  
 QY 463 TKKPAAPTAPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTPKEBAPTTP 522  
 DB 790 AKSPVKEKIDIKPPAEKSPGAKSPA--KSPVKEGAKPPKAPPLDVKSPGKSPGAKSPAEPKSPGAKSPA 848  
 QY 523 TPTTPEE-PATTPKPAAPNTPKPEBAPTTP-----KPEBAPTTP-----KEBAPTTPKET 570  
 DB 849 RP--PEOVKSPAEPKSPGAKSPA--KEEAKTSEKAVKKEVSPVKEVAKKEPPKKEVEEK 904  
 QY 571 APTTPKGAPTTPKEBAPTTPKPAPELAPTTTKEPTSTSD--KPAPTTPKGAPTTP 628  
 DB 905 TLPTPKTEAKSKKDEAPKPAKPKVEKEKTPPEKPKDSTAEKKEEAGEKKKAVASEE 964  
 QY 629 KEPAPTTPKEBAPTTPKGAPTTPKEBAPTTPKPAPELAPTTTKEPTSTSD--KPAPTTPKGAPTTP 688  
 DB 965 ETPAKLVKKEA--KPKETETTTTEAEDTKAKEPS---KPTETEK-----KTTEDKA 1006  
 QY 689 PKETAPTTKPEBAPTTPKPAPELAPTTTKEPTSTSD--KPAPTTPKGAPTTP 748  
 DB 1007 -----KKEEMPAPEK-----KDYKEKTESRKEEK-----KTTEDKA 1034  
 QY 749 EPTPKALENSPEKGVPTTPKPAPELAPTTTKEPTSTSD--KPAPTTPKGAPTTP 808  
 DB 1035 ---PK-NEAVKEDDKSLSEP---SKPTKAKKSSSTDOKESOPPE-----KTTEDKA 1082  
 QY 809 TTEK 813  
 DB 1083 TKGER 1087

RESULT 9  
 TCNA\_TRYCR STANDARD; PRT; 1162 AA.  
 AC P23253;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SIALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).  
 GN TCNA.  
 OS Trypanosoma cruzi.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxID=5693;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SIIVIO X-10/4;  
 RX MEDLINE=91277609; PubMed=1711561;  
 RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,  
 RA Prioli R.P.;  
 RT "The Trypanosoma cruzi neuraminidase contains sequences similar to  
 RT bacterial neuraminidases, YWRD repeats of the low density lipoprotein  
 RT receptor, and type III modules of fibronectin.";  
 RL J. Exp. Med. 174:1179-1191(1991).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=91376547; PubMed=1896773;  
 RA Prioli R.P., Mejia J.S., Aji T., Akawa M., Pereira M.E.A.;  
 RT "Trypanosoma cruzi: localization of neuraminidase on the surface of  
 RT trypanosomes.";  
 RL Trop. Med. Parasitol. 42:146-150(1991).  
 CC -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN  
 CC PARASITE INVASION OF CELLS.  
 CC CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3'- AND 2,8-GLYCOSIDIC  
 CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACYLNEURAMINYLS  
 CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED  
 CC NEURAMINYLS. RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,  
 CC GLYCOLIPIDS OR COLOMINIC ACID.

```

DT 01-MAR-1992 (rel. 21, Last annotation update)
DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN HRGPNT3.
OS Nicotiana tabacum (Common tobacco);
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. XANTHI; TISSUE=Leaf;
RX MEDLINE=90128263; PubMed=2612909;
RA Keller B., Lamb C.J.;
RT "specific expression of a novel cell wall hydroxyproline-rich
RT glycoprotein gene in lateral root initiation.";
RL Genes Dev. 3:1639-1646(1989).
CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC MAIN ROOT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- PPM: EXTENSIN CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X13885; CA32090.1; -.
DR PIR; S06733; S06733.
KW Repeat; Cell wall; glycoprotein; signal; structural protein;
KW Hydroxylation.
FT SIGNAL 1 ?
FT CHAIN 70 620 EXTENSIN.
FT REPEAT 70 73 H-A-P-P.
FT REPEAT 148 151 H-A-P-P.
FT DOMAIN 229 242 2 X 7 AA TANDDEM REPEATS OF T-H-R-H-A-P-P.
FT REPEAT 229 235 1.
FT REPEAT 236 242 2.
FT REPEAT 205 620 CONTAINS THE SER-PRO(4) REPEATS.
FT DOMAIN 499 600 3 X APPROXIMATE TANDDEM REPEATS.
FT FT
SQ SEQUENCE 620 AA: 65406 MW: 641DD2278AB28524 CXC64;

```

Query Match 8.1%; Score 551; DB 1; Length 620;  
 Best Local Similarity 27.2%; Pred. No. 3.1e-16;  
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

```

Db 310 PPTPTPTSPPPPAASPPPTYSPPPTLLPLSSPIYSPPPVYSPPPPSYSPPTLYL 369
QY 515 APTTPEEPPTTPEEPAPTTKKAAP---NTPKEAPTTKKEAPTTKKEAPTTKETA 571
Db 370 PPPPSPPPPSPSPPPPEOSPPPPPAVSPPPLAPPTYSPPPT--YSPPTTYAQP 427
QY 572 PTTTKGATTTTKEAPATTPKKAPKELAPTTTKEPTTTSKRPATTPKGTAPTTKEP 631
Db 428 P-----LPTTSPPPPAVSPPPTTYSPPPTYSPPPAVAQPPPTTYSPPPAVSP 481
QY 632 APTTKEAPATTPKGTATTTTKEAPATTPKKAPKELAPTTTGTSTTSKRPATTTKE 691
Db 482 PPPSPITSPPPQVQYDLPPTTSPPPRRINLPPPHKQF---RPPPTYGQPPSPPTSP 538
QY 692 TAPTTKEAPAT--TPKKAAPTTTTPPTTSEVSTPTTKKPTTTHKSPPDETPELSAE 749
Db 539 PPPQIHSPPPHMQPRTPTTYGQPPSPPTTSAPPPROIHSPPPHKQRPRTPTTYGQF 598
QY 750 PPPKALENSPKRPGVPTTKPATKP 775
Db 599 PSP-----FTTYSPPSPPP 612

```

```

RESULT 8
ID NFH_MOUSE STANDARD; PRT; 1087 AA.
AC P19246; O61959;
DT 01-NOV-1990 (rel. 16, Created)
DT 01-AUG-1992 (rel. 23, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
DE NEFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69121513; PubMed=3220257;
RA Julien J.-P., Cole F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
RT neurofilament subunit.";
RL Gene 68:307-314(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=69089138; PubMed=3145094;
RA Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT "The structure of the largest murine neurofilament protein (NF-H) as
RT revealed by cDNA and genomic sequences.";
RL Brain Res. 464:217-231(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
RA Carden M.J.;
RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PPM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PPM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534

```



FT	DOMAIN	1862	2292	VFED 4.
FT <th>DOMAIN</th> <td>2293</td> <td>2684</td> <th>VFED 5.</th>	DOMAIN	2293	2684	VFED 5.
FT <th>DOMAIN</th> <td>?</td> <td>?</td> <th>EGF-LIKE.</th>	DOMAIN	?	?	EGF-LIKE.
FT <th>CARBOHYD</th> <td>74</td> <td>74</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	74	74	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>403</td> <td>403</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	403	403	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>1023</td> <td>1023</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	1023	1023	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>1099</td> <td>1099</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	1099	1099	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>1618</td> <td>1618</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	1618	1618	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>1737</td> <td>1737</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	1737	1737	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>1832</td> <td>1832</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	1832	1832	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>1878</td> <td>1878</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	1878	1878	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>2136</td> <td>2136</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	2136	2136	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CARBOHYD</th> <td>2505</td> <td>2505</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CARBOHYD	2505	2505	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>CONFLICT</th> <td>2374</td> <td>2379</td> <td>N-LINKED (GLCNAc. . .) (POTENTIAL).</td>	CONFLICT	2374	2379	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT <th>NON_TER</th> <td>2700</td> <td>2700</td> <td>NNQKKA -&gt; RAGEP (IN REF. 1).</td>	NON_TER	2700	2700	NNQKKA -> RAGEP (IN REF. 1).
SO <th>SEQUENCE</th> <td>2700</td> <td>AA: 293013</td> <td>MM: 80E60CC0B12277B1 CRC64:</td>	SEQUENCE	2700	AA: 293013	MM: 80E60CC0B12277B1 CRC64:

```

Query Match      8.28; Score 555.5; DB 1; Length 2700;
Best Local Similarity 32.68; Pred. No. 7e-16;
Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32

QY 123 PINRPSLPNSDISKETSLSLVNNEETVVERKEETTTKQKOSTDGKEKETSAAKTOSIEKT 182
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 445 PVKVLPELPSPVSVS-----TGPSSETGLTENPTLSK-----KPTVSIENK 487
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 SAKLAPTSKVLAKPPTPKAETTTKGALPTTKPEKPTPTTPKPASTTPKEPT-----PTTI 237
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 488 SVTTEKPT-----VPKEKPPIPTKEKPTISYEK---PTIPSEKPNMSEKPTISEKPTLL 539
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 238 KSAPTTKEBPATTTTSAPTTKEBPATTTKEBPATTTKEBPATTTTKEA-PTTTKSAPT 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 540 TEKPTIPSE-KPTIIPSEKPTISEKPTVPEE---PTTPEETTTVVEEVIPIPEKPSIPT 596
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 TPKEBPATTPKKAPPTTPKEBPATTPKEBPPTTPKPEBPATTPKEBPATTPKBPAPKK 356
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 597 --EKPSIPEK---PTISMEETIISIEKPTICEKPTIPEK---PTITTEKSTISPEK- 647
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 357 APTTKEBPATTPKEBPATTTTKEPSTTTKEBPATTTKSAPTTTKEBPATTTKSAPTTPK 416
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 648 -PTTTPPE-KPTITTEKPTISTEKPTIPEK-PTISEKELLIPTKEKLIPT--EKPTIPT 701
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 417 EPSPTTTKEBPATTPKEBPATTPKKPA-----PTTKEBPATTPKEBPATTTKKAPAP 471
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 702 EKPTISTEE-PTTPIEETIISTEKPSIMEKPTLPTEETTTSVETIISTEKLIPT-- 757
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 472 KEBPATTPKEBPATTP-----TPKKLAPPTPEKLA-----PTTPEKBPATTPBELAPTPPEE 521
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 758 EKPTISTEKPTIPTKEKPTISPEKLIPT-EKLIPTKEKPTIPIEETIISTEKL-TIPE 814
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 522 PPTTPEEKPATTPAAANPTKEBPATTPKKPAATTTTKEBPATTTKETAATTPKGAAPT 581
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 815 KPTIPEKKTISTEK-----PTIPE-KPTIPE-----EYTTISTEKLIPT 855
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 582 TLKEBPATTPKKAPKELAPTTTKEPTSTYT-----SKPAPTPKGAAPTTPKEBPATTT 635
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 856 --EKPTIPEKLIPTKEKPTISTEKPTIPTKEKPTIPIEKPPIPT--EKLAALR 911
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 636 PKEBPATTPKGAAPTTLKEBPATTPKKAPKELAPTTTKGSTSTSDKAPATTPKET--A 693
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 912 PPHSPPTL-TGLALVMSPHABSTPMSTV---IGTTTSSSTSGMSP-PNAVYESCAC 966
   | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 694 PTTPEKAPT 703
   | : | | | | |
Db 967 PASCSPAPS 976

```

```

RESULT 5
FPI_MYTED STANDARD: PRT: 875 AA.
ID FPI_MYTED
AC 025460;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
DE PROTEIN 1) (MEPPI) (FRAGMENT).
GN FPI.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91025829; PubMed=1367451;
RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
RT "Structural and functional repetition in a marine mussel adhesive
RT protein.";
RL Biotechnol. Prog. 6:171-177(1990).
RN [2]
RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
RX MEDLINE=83135732; PubMed=6298211;
RA Waite J.H.;
RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
RT hydroxyproline-containing decapeptide in the adhesive protein of the
RT mussel, Mytilus edulis L.";
RL J. Biol. Chem. 258:2911-2915(1983).
CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY
CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
CC (DOPA) DERIVED FROM TYROSINE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X54422; CA38294.1;
CC DR InterPro: IPR002964; Adhesive_Plaq.
CC DR InterPro: IPR002965; P_Rich_extensn.
CC DR PRINTS: PRO1216; ADHESIVE1.
CC DR PRINTS: PRO1217; PRICHEXTENS.
CC Repeat: Hydroxylation.
CC KW NON_TER 1
CC FT DOMAIN 67 870
CC FT P-P-[ST]
CC FT TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
CC SEQUENCE 875 AA; 100412 MW; 6EA85312748CAACE CRC64;

```

```

Query Match 8.2%; Score 556.5; DB 1; Length 875;
Best Local Similarity 28.7%; Pred. No. 2.5e-16;
Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;

```

```

QY 107 TOHNKYSTSPKITTAPKINP-----RP--SLPPNSTSKETSITVKKETTFE 152
DB 1 TKEHPYKPYKTSYSAFYKPPYQPLKKKVDYKPRKTSYPTYG-SKTNYLPLAKKLSYK 59
QY 153 KETTTTKOKSTDG--KEKT-----SAKETOSIKETSAKDLAPTSKYLAKPTPAETTKG 207
DB 60 PIKTTYNAKTNPPYKPKMTYPTTKPKPSYPTTKSKTYTKPKITLYPTTYAKKSY-- 117

```

```

QY 208 PALTPKKEPTPTTKE-----PASTTPKEPTPTTIKSAPTTP-----KEPATTTKSAPT 258
DB 118 PSSYKPKKTPPTTKPKLTYPTTKPKPSYPTTKPKPSYPTTKPKPSYPTTKPKPSY 177
QY 259 PKEPATTTKEPA--PTTPKEPA-----PTTKEPA--PTTKSAAPTTPK--EPAPT 305
DB 178 P-----PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 233
QY 306 PKKPA--PTTPKEPA--PTTKEPT--PTTKEP--APTKEAPATTPKPA--PTAPK 354
DB 234 KAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 293
QY 355 KPA--PTTPKEPA--PTTPKEPA--PTTKEPS--PTTPKEPA--PTTKSAAPTTPKEPA-- 404
DB 294 KPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 353
QY 405 PTTKSAPTTPKEPSPTTKEPAPTTPKEPAPTTPKEPA--PTTKEPA--PTTPKEPA 460
DB 354 PSTYKAKPSY-----PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 409
QY 461 TTTKAPATPAKPEPATPTTKEPATTPKLPPTPEKLAPTTPEKAPATPEELAPTPE 520
DB 410 PPTTKAKPTTKAP-----TYPST-YKAKPSY-----PSYKAKPSYPTTKAKPTYKA 457
QY 521 EPTTPPEEPAPTTPKAAPNTPEKAPATTPKEPATTPKE-----PATTPKEPATPT-- 574
DB 458 KPT-----YPTKYA-----KPSYASAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 506
QY 575 PKGTAPTLKEPATPTTKAPKAPKELAPTTKEPTSTSDKPAPTPKGTAPTT-----PKE 630
DB 507 PKLTLYPTTKA-PKPSYPSYKPKTYPTTKA-----PKISYPTTKAKPSY 551
QY 631 PATTPKEPATTPKGT-----APTTLKEPA--PTTPKPA--PKELAPTTPGK--PTSTSD 682
DB 552 PATYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 611
QY 683 KPAATTPKETAPTTPKEPA--PTTPKPA--PTT-----PEPTPTTSEVSTPTTKE----- 731
DB 612 KAKPSY-----PTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 667
QY 732 PTTTHKSPDSTPELSAEPKALENSPP-----EPGVPTT--KTPA-----ATKPEMTT 780
DB 668 PPT-YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 726
QY 781 AKDPTTERDNTPEPTTAAPKMTKETATTTETTESKITATITQVSTTTOOTPEKIT 840
DB 727 YKAPPTKAKPTYSYKAP--TYKAPTYPTTKA-----DSYPTTKAKPSYPTTKAKPSY 777
QY 841 TLKTTTLAPKVTITTKTITTEIMNKPEETAKPKDRATNSKATTPKPKPTKAPK--P 897
DB 778 TYKSKSTYPSYKPKKTYPTT-----YKPLTYPTTKA-----PKSYPSYKPKITYP 826
QY 898 TSTKPKTTPRVKRPKTTTP-----RKMTSTPELNPSTSR 933
DB 827 STYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 874

```

```

RESULT 6
ZAN_HUMAN STANDARD: PRT: 2700 AA.
ID ZAN_HUMAN
AC 09493; O00218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZONADHESIN (FRAGMENT).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.

```



```

Db      905 YPSQSTETTSSTETTTTTPKTTTTPSTTTTITTTTTCSTGTNSAGETSGCSKPTVT 964
      821 AT--TTQVSTSTTQDTTPFKRTTLTKTTTLPKVT--TKKTTTT-ELMNKEETAKPK 874
      965 TVPTTTTTSVTSSSTTTTITTTTVCSTGTNSAGETSGCSKPTTTTTPCSTSEEA--- 1021
      875 DRATNSATTPKPKP-----TKAPKRTSTKKPKTMPVRKPKPTTP 917
      1022 -----SESTTSPPTTPTTAVTTTAVTTTETSTKSGELTTTFVKNITPTTTLTIAP 1076
      918 PRKMTSTMPETLPTSTRTAEMLQTTTPNQTPNKSLVEVNPKSEDAAGAGETPHMLIR 977
      1077 P--SVTIVNTPTT-ITTTVCSGT-----NSAGETSGCSKPTVTITVPC 1120
      978 HVPEVTPDMQYLPKPPNCGIITNPLSDETNCGKPPDGLTTLNGTLVAFRGHYFW 1037
      1121 STGGEYTTETATLVTTAVTTTAVTTTSTSGTNSA-GKTTTGTYTKSVPTT-----YVT 1173
      1038 MLSPFSPSPARITFEVWGIPSPIDVFTTRC---NCEGKT 1074
      1174 TLAPSAPVTPATN-----AVPTTIT--TECSATVNAAGET 1207

Db      1174 TLAPSAPVTPATN-----AVPTTIT--TECSATVNAAGET 1207

RESULT 4
ID      MUC1_HUMAN
AC      P15942: P13931; P17626: Q14128; Q16442: Q16437; Q9Y4J2;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE      (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE      (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H3A6) (PEAUT-
DE      REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE      DF3).
GN      MUC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Pancreas;
RX      MEDLINE=90368716; PubMed=2394722;
RA      Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT      "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL      J. Biol. Chem. 265:15294-15299(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90202794; PubMed=2318825;
RA      Lichtenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT      "Episialin, a carcinoma-associated mucin, is generated by a
RT      polymorphic gene encoding splice variants with alternative amino
RT      termini.";
RL      J. Biol. Chem. 265:5573-5578(1990).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      MEDLINE=90368715; PubMed=1697589;
RA      Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RT      Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT      "Molecular cloning and expression of human tumor-associated
RT      polymorphic epithelial mucin.";
RL      J. Biol. Chem. 265:15286-15293(1990).
RN      [4]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91097524; PubMed=2268309;
RA      Lancaster C.A., Peat N., Duhig T., Wilson D.,
RT      Taylor-Papadimitriou J., Gendler S.J.;
RT      "Structure and expression of the human polymorphic epithelial mucin
RT      gene: an expressed VNTR unit.";
RL      Biochem. Biophys. Res. Commun. 173:1019-1029(1990).

```

```

RN      [5]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      MEDLINE=90276413; PubMed=2351132;
RA      Wreschner D.H., Hareveni M., Tsarfaty I., Smorodinsky N.,
RA      Horov J., Zaretsky J., Kotkes P., Weiss M., Lache R., Dion A.,
RA      Keydar I.;
RT      "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT      may generate multiple protein forms.";
RL      Eur. J. Biochem. 189:463-473(1990).
RN      [6]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Breast carcinoma;
RX      MEDLINE=90276414; PubMed=2112460;
RA      Hareveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA      Zrlhan S., Weiss M., Green S., Lache R., Keydar I., Wreschner D.H.;
RT      "A transcribed gene, containing a variable number of tandem repeats,
RT      codes for a human epithelial tumor antigen. cDNA cloning, expression
RT      of the transfected gene and over-expression in breast cancer
RT      tissue.";
RL      Eur. J. Biochem. 189:475-486(1990).
RN      [7]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91033045; PubMed=1688329;
RA      Tsarfaty I., Hareveni M., Horev J., Zaretsky J., Weiss M.,
RA      Jeltsch J.M., Garnier J.M., Lache R., Keydar I., Wreschner D.H.;
RT      "Isolation and characterization of an expressed hypervariable gene
RT      coding for a breast-cancer-associated antigen.";
RL      gene 93:313-318(1990).
RN      [8]
RP      PARTIAL SEQUENCE FROM N.A.
RX      MEDLINE=88330762; PubMed=3417635;
RA      Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA      Burchell J.;
RT      "A highly immunogenic region of a human polymorphic epithelial mucin
RT      expressed by carcinomas is made up of tandem repeats.";
RL      J. Biol. Chem. 263:12820-12823(1988).
RN      [9]
RP      SEQUENCE OF 1-169 FROM N.A.
RX      MEDLINE=90088473; PubMed=2597151;
RA      Abe M., Siddiqui J., Kufe D.;
RT      "Sequence analysis of the 5' region of the human DF3 breast
RT      carcinoma-associated antigen gene.";
RL      Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN      [10]
RP      SEQUENCE OF 1-109 FROM N.A.
RC      TISSUE=Thyroid;
RX      MEDLINE=96183746; PubMed=8608966;
RA      Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT      "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT      transcriptase polymerase chain reaction of the MUC1 gene.";
RL      Int. J. Cancer 66:55-59(1996).
RN      [11]
RP      SEQUENCE OF 1-89 FROM N.A.
RC      TISSUE=Lung;
RX      MEDLINE=96181716; PubMed=8604237;
RA      Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA      Lee L.N., Luh K.T., Wu C.W.;
RT      "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT      tissues.";
RL      Oncology 53:118-126(1996).
RN      [12]
RP      SEQUENCE OF 1-46 FROM N.A.
RC      TISSUE=Breast carcinoma;
RA      Bulweila L., Liu Q., Lugmani Y.A., Gomm J.J., Coombes R.C.;
RT      Submitted (OCT-1992) to the EMBL/Genbank/DBJ databases.
RL      -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC      CYTOSKELETON.
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM
CC      IS ALSO PRODUCED.
CC      -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE
CC      SPLICING.
CC      -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL

```

QY 922 -----TSMPELNPSTRIAEAMLOTTRNQTNPNSKLVENPKSEDAAGAGETPMH- 973  
 DB 1454 YLNGYPCGSRPRERNTIRAAAVF-----AKLL-----GADESGAQSASPSYSD 1498  
 QY 974 LLRPH-----VEMPE-----VTPMDLPLPRPNNGI----- 999  
 DB 1499 LADTHMAAAAIKFRATSGLFKGYRPGDTGQITRAEFATVVLHFLTVKKGCEIMSKIA 1558  
 QY 1000 ---LIPMLSDETNINCNG-----KPVGLTTL-----RNGTLVAFRGHYFMMLSPSPSPS 1046  
 DB 1559 TIDISNPKRFD-----CVGHNAOEFIEKLTSLGYISGYPGT-----FKPON 1600  
 QY 1047 PARRTIEWGIPSPIDTFTTRCNCEGKTEFFKD--SQWRFETNDIKD 1091  
 DB 1601 YIKRSESV-----ALINRALEGRPLNGAPKLPDPVNESYNAF-GDIND 1642

RESULT 3  
 AMYL\_YEAST  
 ID AMYL\_YEAST STANDARD: PRT: 1367 AA.  
 AC P08640; P08068;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-  
 GN GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).  
 STAI OR STA2 OR MAL5 OR YIR019C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_Taxid=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5286C / AB972;  
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Cossey T., Dear S., Devlin K., Fraser A.,  
 RA Guelles S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,  
 RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.  
 RX MEDLINE=87194600; PubMed=3106330.  
 RA Yamashita I., Nakamura M., Fukui S.;  
 RT "Gene fusion is a possible mechanism underlying the evolution of  
 RT STA1.",  
 RL J. Bacteriol. 169:2142-2149(1987).  
 RN [3]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RC STRAIN=SPX101-1C;  
 RX MEDLINE=89031230; PubMed=3141213;  
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;  
 RT "Similar short elements in the 5' regions of the STA2 and SGA genes  
 RT from Saccharomyces cerevisiae.",  
 RL FEBS Lett. 239:179-184(1988).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-  
 CC GLUCOSE RESIDUES SUCCESSFULLY FROM NON-REDUCING ENDS OF THE CHAINS  
 CC WITH RELEASE OF BETA-D-GLUCOSE.  
 CC -1- SIMILARITY: TO S.POMBE SPC215.13.  
 CC -1- SIMILARITY: TO S.POMBE SPC285.13C.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: 238061; CAA86176.1; -  
 DR EMBL: M16164; AAA35014.1; -

DR EMBL: M16165; AAA35015.1; -  
 DR EMBL: X13857; CAA32069.1; -  
 DR PIR: B26877; B26877.  
 DR PIR: A26877; A26877.  
 DR PIR: S48478; S48478.  
 DR SGD: S0001458; MOC1.  
 KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1367  
 FT DOMAIN 210 1367  
 FT CARBOHD 817 817  
 FT CARBOHD 874 874  
 FT SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CnC64;

Query Match 11.7%; Score 798.5; DB 1; Length 1367;  
 Best Local Similarity 27.9%; Pred. No. 5,9e-26;  
 Matches 313; Conservative 108; Mismatches 523; Indels 177; Gaps 50;

QY 44 DDAQCKKKYDK--CCPDYESFCAEVK-DNKKNRKKKPKRPVYDE--AGSGLDNGDF 97  
 DB 174 DLSTGCNNYDNOGHSQTDPEGFYWNIDCANNCGCTKSTTSSSTSTSSSTTSSSTT 233  
 QY 98 KVTTPDSTQHNKSVTSKRTTKAKLPN-----RSLPRNSDTSKETSLVKNKETT 149  
 DB 234 TSSSTSSSTTSSSTSSSTSSSTSSSTTAAPRTTTSCTKEKRPRTTTSCTKEKRPRTT 293  
 QY 150 VETKEITTTTKQNTSDKEKTSKETSKEKTSKETSANDALPTSVLAKPKPKAKET--TKG 207  
 DB 294 PCTKKKTTSK-TCY--KTTTPVPTPS--SSTESSAPV-----PTSSSTSSSA 342  
 QY 208 PALTPKE-----PPTTPKEPASTTPKEPPTTKSAP-----TPKEPAPPTTKSAP 257  
 DB 343 PVTSSTSSAPVPPTSSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVPPTSSST 402  
 QY 258 TPKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKAPPTTKAPPTTKAPPTTKEP 317  
 DB 403 TESSAPVTS-----TTSSAPVTS-----STSSAPVTS-----SSTSSAPVTSST 451  
 QY 318 APPTPKPPTP--TPKEPAPPTTKEPAPPTTKEPAPPTTKAPPTTKAPPTTKEPAP 370  
 DB 452 ESSAPVPPTSSSTSSAPVTS--SSTSSAPV--VTPSSSTSSAPVTSSTSS 507  
 QY 371 EPAPPTTKPPTTKEPAPPTTKSAPPTTKEPAP--TTKSAPTTKPPTTKPPTTKP 426  
 DB 508 SSAPVPPTPS--SSTSSAPAPPTSSSTSSAPVTSSTSSAPVTPPTSSSTSS 566  
 QY 427 AP--TPKEPAPPT--TPKPAPTTKEPAPPTTKEPAPPTTKKAPPTTKEPAPPTTK 481  
 DB 567 TPVTSSTSSAPVPPTSSSTSSAPVPT-----PSSSTSSAPAPVPPTSSSTSS 622  
 QY 482 TAPPTTKKLTPTTPEKLAPTTPEKAPPTTPEELAPT--TPEPTPTTPEEPAP--TTP 535  
 DB 623 SAPVT--SSTSSAP--VPTSSSTSSAPVPPTSSSTSSAPVPTSSSTTE 677  
 QY 536 KAAPPTKEPAPPTTKEPAPPTTKEPAPPTTKEPAPPTTKGAPPTTKEPAPPTTKP 594  
 DB 678 SSAPVT--SSTSSAPVT--SSTSSAPVT--SSTSSAPVT--SSTSSAPVT--SSTSS 729  
 QY 595 AKKELAPPTTKKPTSTSDKAPPTPKGTAPTTKEPAP--TPKEPAPPT--TPGTAAPT 650  
 DB 730 TESSAPVPTPS--SSTSSAPVTSSTSSAPVPTSSSTSSAPVPTSSSTSSAPVPTSS 788  
 QY 651 TLKEPAPPTPKKAPKELAPTTTKGPTSTSDKAPPTTKETAAPTTKEPAP--TTP 705  
 DB 789 SSAPVPTSSSTSSAPVPTSSSNTTSSAPSSSTSSSTSSAPVPTSSSTTE 848  
 QY 706 KKPAPPTTPTPTTSEVSTPT--TTKEPTTIHKSPTDSPTELSDEPTKALESPK 760  
 DB 849 SSAPVSSSTSSAPVPTSSSNTTSSAPSSI--PSSSTSSSTGT--TTPSSSK 904  
 QY 761 EPVPTTTPAATKPEMTTKAKDKTTERDLATTPETTTAAPKMKETATTEKTESKIT 820

```

QY 565 T-TPKETAPTPKGAAPTLKEBAPTPPKKPAKELAPTTTKETSTSDKAP--TPKGG 622
DB 1958 TPTGOTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2017
QY 623 TAPTTKEBAPTPPKBAPTPPKGAPTLKEBAPTPPKKPAKELAPTTTKGPTSTSD 682
DB 2018 TTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2076
QY 683 KPAAP-TTPKETAPTPKKEBAPTPPKKPAAPTPPTPTPTPTPTPTPTPTPTPTPTPT 741
DB 2077 TQPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2132
QY 742 STEPLASPPPKALENSPKAP-----GVPT--TKTPAATKPEMTTAAKDKTERDL 790
DB 2133 TTTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2192
QY 791 R---TPP--ETTPAAPKMT--KETATTEKTESKITAATTOVSTTODTTPFKITL 842
DB 2193 QTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2248
QY 843 KTTTLAKRVTTT-KKLTITTEIMKPEETAKPKDRATNSKATPPKPKAP-KKPTST 900
DB 2249 -TTTTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2307
QY 901 KKPTMPKPVKPKPTPTPTPKKMTSTMPKELNP--TSRIAEMALQTTTR-PNOTPSKLVEY 956
DB 2308 QTPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2367
QY 957 NPKSEDAGAGETPPHMLLRPHVMEPYTP 986
DB 2368 TPTPTPTG---TQTP--TTTPTTTTTTPT 2392

RESULT 2
SLP1_CLOTM STANDARD; PRT: 1664 AA.
ID SLP1_CLOTM
AC 006852;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE PROTEIN 1).
GN OLPB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=9320931; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein Cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RT J. Bacteriol. 175:1891-1899(1993).
RL J. Bacteriol. 175:1891-1899(1993).
CC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOMOLOGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC EMBL: X67506; CAA47841.1;
CC InterPro: IPR001119; SLH.
CC Pfam: PF00395; SLH. 3.
CC PROSITE: PS01072; SLH_DOMAININ. 2.

```

```

KW Cell wall; S-layer; Signal; Repeat.
FT SIGNAL 1 28
FT CHAIN 29 1664
FT DOMAIN 36 763
FT REPEAT 36 191
FT REPEAT 207 363
FT REPEAT 409 565
FT REPEAT 607 763
FT DOMAIN 771 1377
FT DOMAIN 1378 1449
FT DOMAIN 1453 1494
FT DOMAIN 1495 1565
FT DOMAIN 1566 1625
FT DOMAIN 1626 1646
SQ DOMAIN 1664 AA; 178194 MW; 5F39695BA5FE/4B CRC64;

Query Match 13.9%; Score 950; DB 1; Length 1664;
Best Local Similarity 31.0%; Pred. No. 5,1e-32;
Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 193 VLAQPP-RAETTKGAPALTPEKEPTPTPKBAPSTTPKEPTPTTKSAPTPKKEBAPT 251
DB 758 VVIQAPAKASDEPIPTDPSDEPTPS-----DEPTPS--DEPTPSDEPTPTDTP 804
QY 252 TKAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPK 307
DB 805 EPPPTSEEPPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 864
QY 308 KPAAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPKBAPTPPK 366
DB 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--DEPTPSDEPTPEPIPTDPSDEPTPSDEPT 919
QY 367 TTKKEPA--PTTKESAPT--TKKEBAPTTS--APTTPKEAPT--TTKSAPTTKEPS 419
DB 920 SDEPTPSDEPTPSDEPTPEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 979
QY 420 ----PTTKBAPTP--TKBAPTPPKKAPTPPKBAPTPPKBAPTPPKBAPTPPKKAPTPK 474
DB 980 PSDEPTPSDEPTPEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1039
QY 475 APT--TPKETAPTPPKKLTPTPEKLAPTTPKEKAPT--TPEELAPTPEEPTPT--TP 527
DB 1040 TPSDEPTPSDEPT--TPEEPIPTDPSDEPTPSDEPTPSDEPTPSD--EPTSDSEPTSEPT 1094
QY 528 EEPAPTPPKAAPTTPKEBAPTPKEBAPTP--PTTKBAPTP--TKENAAPTTPKGAPTTLK 584
DB 1095 EEPITDTPPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPEPIPTDPSDEPTPSD 1154
QY 585 EAPPTPKKPAKELAPTTTKEPT--STTSKPAATTTKGAAPTTPKKEBAPTPKKEBAPT 643
DB 1155 EPTPS--DEPTPSD--EPTPSDEPTPEPIPTDPSDEPTPSDEPTPS--DEPTPS-- 1208
QY 644 PKGTAPTLKBPAPT--TPKKPAKELAPTTTKGPTSTTSKAPAPT--TPKETAPTPKKE 699
DB 1209 ---DEPTPSDEPTSETEPEPI-----PTDPSDEPTPSDEPTPSD--EPTPSDE 1253
QY 700 PAPTPPKKPAAPT--TPETPPTTSEVSTPTTKKEPTTIHKSDESTPELSAPPTKALENS 758
DB 1254 PTPS--DEPTPESEPEPIPTDPSDEPTPSDEPT--PSDEPTPS--SDEPTPSDEPTPT 1305
QY 759 PKEPVPPTTKTPAAT--KPEMTTAKDKTERDLATPEETTAAPAKMTKETATTEKTE 816
DB 1306 SDEPTPSDEPTSETEPEPIPTDPSDEPTPSD---EPTSDSEPTPSDEPTPSDEPTPS 1361
QY 817 SKITATTTQVSTTQDTPTPEKITTLKTTTLAKVTTTKKTTTTEIMNKKEBETAKPKDR 876
DB 1362 DEPTPSDEPTPEPEPT-----TTTPTPTST-----PTSG 1396
QY 877 ATNSKAT-----TPKQKPTKAP--KKPTSTKPKTPKPKKPTTPPKKM----- 921
DB 1397 SGSGSGSGGGGGGGGTVPSTPTPTSKPTSTPAP--TEIEPTPSDVGAIGENRA 1453

```

FT	CARBOHYD	1787	1787
FT	CARBOHYD	1820	1820

511 PEEBAPITPEEPPIPT-JPEEBAPITPKAAAPNTPEKAPITTPKEP-----ADTPKEBAP 564

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:21 ; Search time 62.75 Seconds  
(without alignments)  
742.061 Million cell updates/sec

Title: AA3  
Perfect score: 6814  
Sequence: 1 MAKKTLPIYLLLLSVFVIO.....AAATITRSQGLSKWYNCP 1270

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

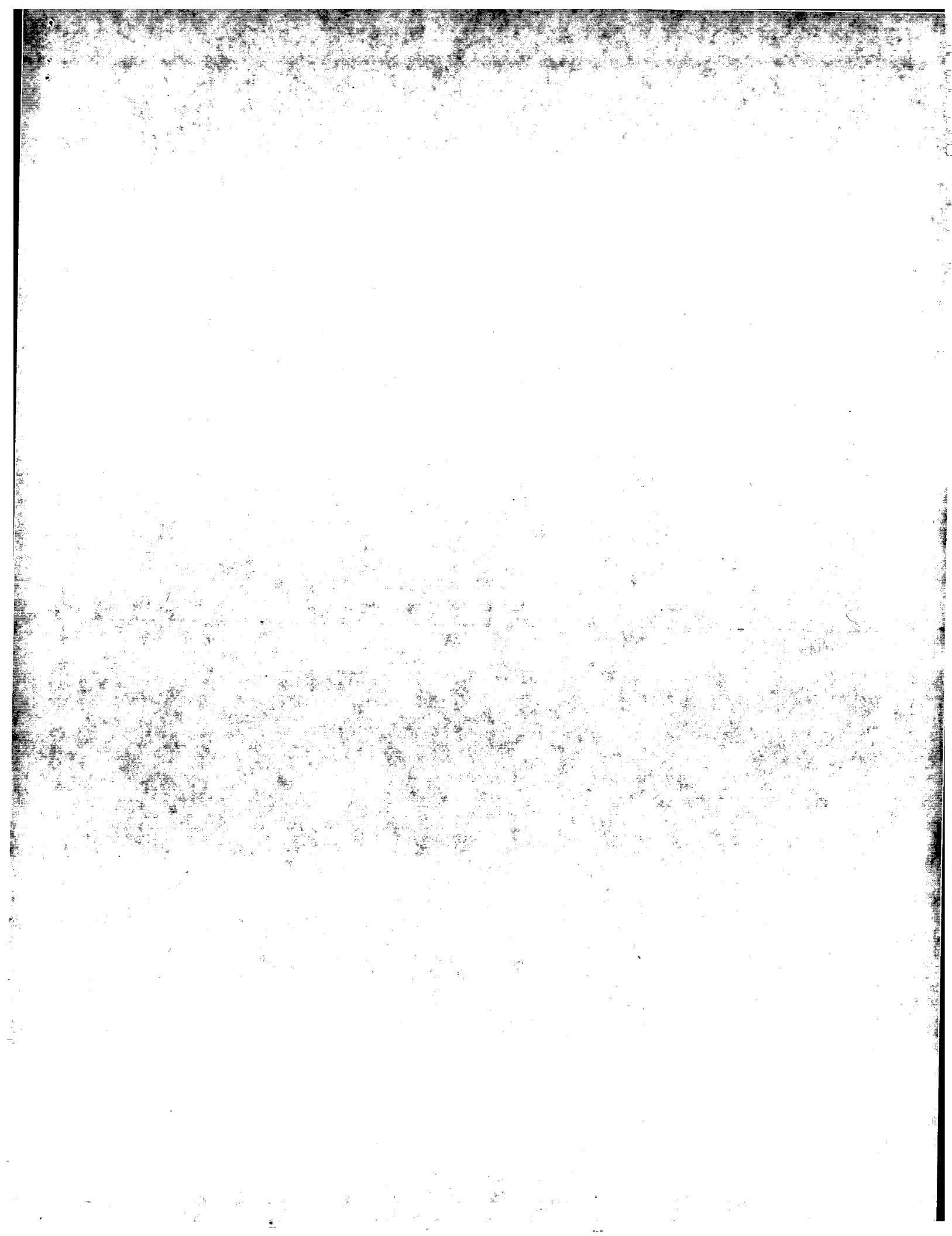
Result No.	Score	Query Match	Length DB	ID	Description
1	1171	17.2	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.9	1664	1 SLPI_CLOTH	Q06852 clostridium
3	798.5	11.7	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	9.6	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	8.2	875	1 FPL_MYTED	Q25460 mytilus edu
6	555.5	8.2	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	551	8.1	620	1 EXTN_TOBAC	P13983 nicotiana t
8	533	7.8	1087	1 NFN_MOUSE	P13946 mus musculu
9	530.5	7.8	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.8	865	1 CPN_DROME	Q02910 drosophila
11	518.5	7.6	872	1 FPL_MYTCO	Q25434 mytilus cor
12	509.5	7.5	662	1 MUC1_XENLA	Q05049 xenopus lae
13	503.5	7.4	1970	1 RPBI_HUMAN	P24928 homo sapien
14	498.5	7.3	1970	1 RPBI_MOUSE	P08775 mus musculu
15	497	7.3	831	1 NFN_RAT	P16844 rattus norv
16	493.5	7.2	467	1 RPBI_CRIGR	P11414 cricetus
17	488.5	7.2	826	1 SSP2_PLAYO	Q01443 plasmodium
18	488	7.2	1020	1 NFN_HUMAN	P12036 homo sapien
19	475.5	7.0	267	1 EXTN_MAIZE	P14918 zea mays (m
20	472	6.9	2142	1 BAT2_HUMAN	P48634 homo sapien
21	471.5	6.9	5376	1 ZAN_MOUSE	O88799 mus musculu
22	468.5	6.9	634	1 HMP1_CANAL	P46593 candida alb
23	467	6.9	817	1 VRP1_YEAST	P37370 saccharomyc
24	455	6.7	1161	1 YJ9P_YEAST	P47179 saccharomyc
25	454	6.7	797	1 VGLX_HSEB	P28968 equine heip
26	448.5	6.6	670	1 VGLX_HSEB	Q00130 ictalurid h
27	442.5	6.5	751	1 FPL_MYTGA	Q27409 mytilus gal
28	439.5	6.4	1083	1 T2D3_HUMAN	O00268 homo sapien
29	432.5	6.3	439	1 XPT_XENLA	P17437 xenopus lae
30	432	6.3	3164	1 TEGU_HSV11	P10200 herpes simp
31	430.5	6.3	3421	1 TEGU_HSEB	P28955 equine heip
32	426	6.3	2715	1 TRX2_HUMAN	O9um66 homo sapien
33	424.5	6.2	1125	1 MAP4_MOUSE	P27546 mus musculu

34	424	6.2	307	1	SGS3_DROME	P02840 drosophila
35	424	6.2	1251	1	YU3_CAEEL	O09550 caenorhabdi
36	422.5	6.2	1794	1	YAV1_SCHPO	O10172 schizosach
37	422	6.2	2774	1	MAPA_RAT	P34926 rattus norv
38	421	6.2	1229	1	N121_HUMAN	Q092n3 homo sapien
39	420	6.2	1411	1	TCOF_HUMAN	O13428 homo sapien
40	419.5	6.2	2476	1	ZAN_PIG	Q28983 sus scrofa
41	419	6.1	1185	1	DRPL_HUMAN	P54259 homo sapien
42	411.5	6.0	907	1	DRPL_DBV	P03200 epstein-bar
43	410	6.0	1183	1	DRPL_RAT	P54258 rattus norv
44	410	6.0	2517	1	NCR2_HUMAN	O9y618 h nuclear r
45	407.5	6.0	3256	1	K167_HUMAN	P46013 homo sapien

## ALIGNMENTS

RESULT	ID	SEQUENCE	STANDARD	PRT	AA
1	MUC2_HUMAN	Q02817; Q14878; 01-JUN-1994 (Rel. 29, Created)			
AC	002817; Q14878; 01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).				
GN	MUC2 OR SMC.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RC	MEDLINE=94132002; PubMed=8300571;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;				
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.				
RT	Identification of the amino terminus and overall sequence similarity				
RT	to prepro-von Willebrand factor.";				
RL	J. Biol. Chem. 269:2440-2446(1994).				
RN	[2]				
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.				
RC	TISSUE=Colon;				
RX	MEDLINE=93016075; PubMed=1400449;				
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,				
RA	Kim Y.S.;				
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located				
RT	both upstream and downstream of its central repetitive region.";				
RL	J. Biol. Chem. 267:21375-21383(1992).				
RN	[3]				
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.				
RX	MEDLINE=91358717; PubMed=1885763;				
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,				
RA	Petersen G.M., Kim Y.S.;				
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays				
RT	and polymorphism.";				
RL	J. Clin. Invest. 88:1005-1013(1991).				
CC	- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND				
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A				
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS				
CC	AGENTS AT MUCOSAL SURFACES.				
CC	- SUBUNIT: MULTIMERIC.				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,				
CC	BRONCHUS, CERVIX AND GALL BLADDER.				
CC	- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR				
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).				
CC	- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND				
CC	VARIABLE AMONG DIFFERENT ALLELES.				
CC	- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT				
CC	OF SILKWORM HEMOCYTIN.				
CC	- SIMILARITY: CONTAINS 2 WWFC DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).				





Mon Apr 29 08:35:18 2002

aa3.rpr

Page 13

OY 800 APRMREIATTEKTESKITATTOVSTTQDTTP 836  
DB 704 PPVKPPPVQVPPPTYSPPIKPPPVQVPPPTTSPSP 740

Search completed: April 26, 2002, 16:31:52  
job time: 707 sec